PCT

WORLD INTELLECTUAL PROPERTY ORGANIZATION
International Bureau

INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(51) International Patent Classification 6 :
C07K 14/00, C12Q 1/68, G01N 33/53, C12N 15/85, A61K 48/00

A2

(11) International Publication Number: WO 98/45322

(43) International Publication Date: 15 October 1998 (15.10.98)

(21) International Application Number: PCT/IB98/00705

(22) International Filing Date: 2 April 1998 (02.04.98)

(30) Priority Data:
60/043,163 10 April 1997 (10.04.97) US

(71) Applicants *(for all designated States except US)*: ROYAL NETHERLANDS ACADEMY OF ARTS AND SCIENCES [NL/NL]; P.O. Box 19121, NL–1000 GC Amsterdam (NL). ERASMUS UNIVERSITY ROTTERDAM [NL/NL]; DR Molewaterplein, Burg Ondlaan 50, Postbox 1738, NL–3000 DR Rotterdam (NL). UNIVERSITY OF UTRECHT [NL/NL]; Universiteitsweg 100, De Uithof, NL–3000 DR Rotterdam (NL).

(72) Inventors; and
(75) Inventors/Applicants *(for US only)*: VAN LEEUWEN, Frederik, W. [NL/NL]; Kerkweg 37, NL–3063 CL Maarssen (NL). GROSVELD, Franklin, G. [NL/NL]; Jacobusvan Vessemsingel 9, NL–3065 NH Rotterdam (NL). BURBACH, Johannes, Peter, Henri [NL/NL]; Koperslagershoek 9, NL–3981 SB Bunnik (NL).

(74) Agents: MASCHIO, Antonio et al.; D. Young & Co., 21 New Fetter Lane, London EC4A 1DA (GB).

(81) Designated States: AL, AM, AT, AU, AZ, BA, BB, BG, BR, BY, CA, CH, CN, CU, CZ, DE, DK, EE, ES, FI, GB, GE, GH, HU, IL, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MD, MG, MK, MN, MW, MX, NO, NZ, PL, PT, RO, RU, SD, SE, SG, SI, SK, SL, TJ, TM, TR, TT, UA, UG, US, UZ, VN, YU, ZW, ARIPO patent (GH, GM, KE, LS, MW, SD, SZ, UG, ZW), Eurasian patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European patent (AT, BE, CH, CY, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, ML, MR, NE, SN, TD, TG).

Published
Without international search report and to be republished upon receipt of that report.

(54) Title: DIAGNOSIS METHOD AND REAGENTS

(57) Abstract

The invention relates to methods and reagents for the diagnosis and treatment of a disease caused by or associated with an RNA molecule having a transcript mutation giving rise to a frameshift mutation. The diagnostic methods include the steps of providing a body fluid or tissue sample from a patient; and analyzing the sample for the presence of an RNA molecule having a frameshift mutation or a protein encoded thereby, wherein the presence of the mutated RNA molecule or encoded protein is indicative of the disease. The therapeutic treatments include administering substances which selectively eliminate mutate RNA molecule from the cell.

DIAGNOSIS METHOD AND REAGENTS

BACKGROUND OF THE INVENTION

The invention encompasses methods and reagents for the diagnosis of a disease caused by or associated with a transcript mutation giving rise to a frameshift mutation within an RNA molecule. The methods include the steps of providing a body fluid or tissue sample from a patient; and analyzing the sample for the presence of an RNA molecule having a frameshift mutation or a protein encoded thereby, wherein the presence of the mutated RNA or encoded protein is indicative of the disease.

It is an object of the present invention to provide methods and assays for detection and/or treatment of diseases involving transcript mutations, particularly those diseases relating to aging, wherein the probability of having the disease increases with the age of the patient. The invention contemplates detection and/or treatment of those age-related diseases which are due to mutations occurring in the RNA of cells. If the mutations are not corrected, the disease may result.

Another object of the invention is to treat diseases identified according to the invention, by providing to a patient afflicted with the disease or having a propensity to develop the disease, a corrective agent such as an enzyme or oligonucleotide.

Yet another object of the invention is to provide a method for identifying age-related diseases by correlating nucleotide sequence mutation hotspots with the disease.

Other objects of the invention relate to identification, detection and treatment of age-related diseases including cancers (especially non-hereditary cancers) and neurodegenerative diseases, such as Alzheimer's Disease (AD), Downs' syndrome, frontal lobe dementia (Pick's Disease), progressive supranuclear palsy (PSP) and other diseases with abundant tau-positive filamentous lesions (such as Corticobasal degeneration, Dementia pugilistica, Dementia with tangles only, Dementia with tangles and calcification, Frontotemporal dementias with Parkinsonism linked to chromosome 17, Gertsmann-Strässler-Scheinker disease with tangles, Myotonic dystrophy, Niemann-Pick disease type C, Parkinsonism-dementia complex of Guam,

Postencephalitic Parkinsonism and Subacute sclerosing panencephalitis), Parkinson's Disease (PD), amyotrophic lateral sclerosis, Huntington's Disease, multiple sclerosis, dementia with Lewy bodies, multisystem atrophy, other inclusion body diseases associated with ubiquitin (such as Alexander's disease, Alcoholic liver disease, lichen amyloidosis, and during aging Marinesco bodies and Hyaline inclusions), diabetes mellitus type II and other degenerative diseases, such as cardiovascular diseases and rheumatoid arthritis. Early disease diagnosis is imporfant for effective treatment.

Alzheimer's Disease is in most cases a disease which is related to aging. AD is characterized by atrophy of nerve cells in the cerebral cortex, subcortical areas, and hippocampus and the presence of plaques, dystrophic neurites, neuropil threads and neurofibrillary tangles. In most cases, it is not known whether AD is caused by a genetic abnormality or by environmental factors, or both. The pathogenic mutation is unknown.

Another object of the invention is to provide a diagnostic test for AD which enables definitive diagnosis of AD in living patients. Furthermore, as AD is a progressive disease, it is desirable to diagnose AD as early as possible so that preventative action may be taken.

A number of diagnostic methods have been previously suggested for AD diagnosis, most of which have focused on the β-amyloid precursor protein. See for example U.S. Patents 4,666,829, 4,816,416 and 4,933,159. However, β-amyloid deposits have been found in individuals, especially aged persons, who have not shown signs of dementia (See J. Biol. Chem., 265, pp 15977, 1990; and Tables 3-5). Diagnostic tests based on the β-amyloid protein have therefore been shown to lack specificity for AD.

In U.S. Patent 4,727,041 a diagnostic test for AD is described based on measuring levels of somatotropin and somatomedin-C in blood sera following administration of an L-dopa proactive test.

In International patent application WO 94/02851, a method is described for identifying AD by the use of antibodies having affinity for paired helical filaments in order to determine the levels of paired helical filaments in cerebral spinal fluid. The

presence of paired helical filaments is alleged to be indicative of AD.

Other diagnostic methods are based on the identification of "disease specific marker proteins" in the cerebrospinal fluid. In International patent application WO 95/05604, for example, five disease specific proteins are identified by their molecular weights. However, the specific identity of the proteins is unknown and their specific relationship to the pathogenesis of AD is also unknown. The five "disease specific marker proteins" may therefore be present as a result of a more fundamental cellular or biochemical change.

Another object of the invention is to provide for detection of AD preferably early on in the disease state. It is desirable to detect a protein or substance which is either directly responsible for the disease or is involved early on in the pathogenesis of the disease, or if not involved is nevertheless generated directly or indirectly by the mechanism causing the disease. Such a protein or substance may be the "causative" agent to the disease or may be "associated with" the disease state in the sense of being diagnostic of the disease state.

Recently, Sherrington *et al.* in Nature, 375, pp 254-260, 1995, identified a gene on chromosome 14 bearing missense mutations which are associated with up to 33% of autosomal dominant early onset AD cases (Table 1). A missense mutation involves a nucleotide substitution, usually a single nucleotide substitution, which results in an amino acid substitution at the corresponding codon. The missense mutations disclosed in Sherrington *et al.* are predicted to change the encoded amino acid at the following positions (numbering from the first putative initiation codon) Met to Leu at codon 146, His to Arg at codon 163, Ala to Glu at codon 246, Leu to Val at codon 286, Cys to Tyr at codon 410. It has been proposed that these mutations may be useful in identifying early onset AD. As stated earlier, the majority of AD cases are late onset (after 65 years of age; Table 1) and it is therefore still a problem to identify the majority of individuals having AD, particularly late onset AD.

There is no indication that these diseases occur at the RNA level and not at the DNA level. Accordingly, the prior art methods of detection are for mutated DNA or for a protein encoded by the mutated DNA, and will not give an indication of the

presence of a transcript mutation in an RNA molecule.

Presently, there are a number of substances which are alleged to be useful in the treatment of AD. However, so far only limited success has been achieved with these substances. Methods for effectively treating and/or preventing AD are still required (see Allen and Burns, Journal of Psychopharmacology, 9, pp 43-56, 1995).

SUMMARY OF THE INVENTION

The present invention is based on the observation that an RNA molecule containing a frameshift mutation and encoding a corresponding mutant protein are correlated with the presence of a disease.

According to the present invention there is provided a method for the diagnosis of a disease caused by or associated with an RNA molecule having one or more mutations giving rise to a frameshift mutation comprising: i. providing a biological sample, such as a body fluid or tissue sample, from a patient; and ii. analyzing the sample for the presence of an RNA molecule having a frameshift mutation or a mutant protein encoded thereby, wherein the presence of the mutated RNA or mutant protein is indicative of the disease.

A "mutant" protein is a polypeptide encoded by a mutant mRNA at least a part of which is in a reading frame that is shifted relative to the initiation start codon from that of the native or wild-type reading frame, and thus will include any protein having an aberrant carboxy terminal portion which is encoded by the +1 or +2 reading frame of the wild type gene sequence. Thus, the mutant protein will include a hybrid wild-type/nonsense protein having an amino terminal amino acid sequence that is encoded by the wild type (O) reading frame and a carboxy terminal amino acid sequence that is encoded by the +1 or +2 reading frame, and thus the nonsense portion of the mutant protein. The cross-over point between the wild type and nonsense amino acid sequences is the codon containing the frameshift mutation.

The invention is based on the discovery of the presence of such a mutant protein or an accumulation of more than one mutant protein in a tissue from a diseased individual, and also on identification of the mutant protein as indicative of the disease.

The invention is also based on the discovery that the mutation that gives rise to the mutant protein occurs at the RNA level and not at the DNA level.

The phrase "caused by or associated with" refers to an RNA molecule which is either fully or partly responsible for the disease, or an RNA molecule which is not responsible for the disease but is associated with the diseased state in the sense that it is diagnostic of the diseased state.

A disease caused by or associated with at least one RNA molecule having one or more mutations giving rise to a frameshift mutation can be any disease including non-hereditary cancers, neurodegenerative diseases such as Alzheimer's Disease (AD); Downs' syndrome; frontal lobe dementia (Pick's Disease); progressive supranuclear palsy (PSP) and other diseases with abundant tau-positive filamentous lesions such as Corticobasal degeneration, Dementia pugilistica, Dementia with tangles only, Dementia with tangles and calcification, Frontotemporal dementias with Parkinsonism linked to chromosome 17, Gertsmann-Strässler-Scheinker disease with tangles, Myotonic dystrophy, Niemann-Pick disease type C, Parkinsonism-dementia complex of Guam, Postencephalitic Parkinsonism and Subacute sclerosing panencephalitis; Parkinson's Disease (PD) amyotrophic lateral sclerosis; Huntington's Disease; multiple sclerosis; dementia with Lewy bodies, multisystem atrophy and other inclusion body diseases associated with ubiquitin such as Alexander's disease, Alcoholic liver disease, lichen amyloidosis, and during aging Marinesco bodies and Hyaline inclusions; and other degenerative diseases such as cardiovascular diseases, rheumatoid arthritis and Diabetes mellitus type II. Cancers treatable according to the invention include, but are not limited to, Hodgkin's disease, acute and chronic lymphocytic leukemias, multiple myeloma, breast, ovary, lung, and stomach or bladder cancers.

An RNA molecule having a transcript mutation which leads to a frameshift mutation, and herein referred to as the "mutant RNA", can be any RNA molecule having at least one transcript mutation which leads to a frameshift mutation. The RNA molecule may be any RNA molecule including primary transcripts and messenger RNA (mRNA).

The term "transcript mutation" refers to a mutation which occurs at the RNA level but does not occur in the DNA from which the RNA was transcribed. In order to identify transcript mutations a comparison between the RNA and the DNA from which the RNA was transcribed has to be made.

A "frameshift mutation" refers to a deletion or insertion of one or more nucleotides within an open reading frame, for example, a single nucleotide or dinucleotide deletion or insertion, such that the reading frame of the coding region is shifted by one or two nucleotides. Preferably, the frameshift mutation is a nucleotide or dinucleotide deletion leading to a + 1 or +2 frameshift mutation. However, any number of nucleotide deletions can occur provided a frameshift mutation results. Alternatively, the insertion of one or more nucleotides may give rise to a frameshift and such mutations also form part of the present invention.

Other genetic modifications which give rise to a frameshift also form part of the present invention, such as a change in the nucleotide sequence which leads to translation initiation from a different position or a mutation outside a coding region, such as within an Intron (if the RNA molecule is a primary transcript), or a 5' or 3' untranslated region, which mutation may result in mis-translation and production of a mutant protein.

It is preferred that the mutation is a nucleotide and more preferably a dinucleotide deletion or insertion associated with the nucleotide sequence GAGA or its complementary sequence CTCT of the RNA molecule; especially preferred frameshift mutations are associated with the nucleotide sequence of the RNA comprising GAGAX or CTCTX, where X is one of G, A, U or C, the preferred motifs being GAGAG, GAGAC, GAGAT, and GAGAA as well as CTCTC, CTCTG, CTCTA and CTCTT. As used herein, the term "GAGA mutation" may refer to either a single nucleotide insertion or deletion or a dinucleotide insertion or deletion within the GAGA or CTCT motif itself or adjacent to (5'- or 3'-terminal to-, and within 5-10 nucleotides of-) the GAGA or CTCT motif.

Preferably, the dinucleotide deletion is a GA deletion within the GAGA motif or a GT deletion immediately following (i.e., within 10 nucleotides 3' of) a

GAGA motif and a CT deletion in the CTCT motif. It is further preferred that the mutant RNA has one or two dinucleotide deletions associated with a GAGA, GAGAC, GAGAG, GAGAT or GAGAA, or with a CTCT, CTCTG, CTCTC, CTCTA or CTCTT, leading to a + 1 or + 2 frameshift mutation respectively.

In a preferred embodiment of the invention, the transcript mutations occur in RNA molecules of the neuronal system, where the disease is a neurodegenerative disease.

The "neuronal system" is defined as any cells, RNA molecules, proteins or substances relating to or forming part of the neuronal system such as nerve cells, glial cells, proteins including Tau, β amyloid precursor protein, ubiquitin B, apolipoprotein E4, neurofilament proteins and microtubule associated protein II, presenilin I, presenilin II, Big Tau, glial fibrillary acidic protein (GFAP), Human P53 cellular tumor antigen, human B-cell leukemia/lymphoma 2 (BCL-2) protooncogene, semaphorins human homolog of yeast up-frameshift protein 1 (HUPF-I), Human Motility Group Protein (HMG), neuron specific protein A (NSP-A) and the RNA molecules encoding the proteins.

Where the disease is a neurodegenerative disease, especially AD, the preferred mutant RNA molecules of the present invention are those encoding the β amyloid precursor protein, the Tau protein, ubiquitin, apolipoprotein-E_4 (Apo-E_4), microtubule associated protein II (MAP 2), the neurofilament proteins, presenilin I, presenilin II, Big Tau, GFAP, P53, BCL2, HUPF-I, HMG and NSP-A, having a deletion, insertion or other modification leading to a frameshift mutation. The most preferred mutant RNA molecules of the present invention are those encoding β amyloid precursor protein, ubiquitin B, MAP 2, the neurofilament proteins, presenilin I, presenilin II, Big Tau, GFAP, P53, bcl2 and HUPF-I, which have a frameshift mutation.

It is preferred that the mutation is a GA or a GT dinucleotide deletion associated with (within or within 10 nucleotides 5' or 3' of) a GAGA or GAGAX sequence leading to a frameshift mutation or a CT or CA dinucleotide deletion associated with (within or within 10 nucleotides 5' or 3' of) a CTCT or CTCTX

sequence leading to a frameshift mutation. It is further preferred that the mutant RNA molecule has one or two GA or GT deletions or one or two CT or CA deletions, each associated with a GAGA or CTCT sequence or similar motif, leading to a + 1 or + 2 frameshift mutation, respectively.

The term "mutant protein" as used herein is defined as the protein encoded by the mutant RNA molecule of the present invention.

It is preferred that the methods of the present invention are for the diagnosis of a disease caused by or associated with at least one RNA molecule having one or more transcript mutations giving rise to a frameshift mutation. A preferred disease for diagnosis by the present invention is AD, except the early onset AD cases found to be linked to chromosome 1, 14 and 21. It is further preferred that the methods of the present invention are for the diagnosis of young and late onset AD, especially non-familial or "sporadic" late onset AD cases.

As used herein, "biological sample" refers to a body fluid or body tissue which contains proteins and/or cells from which nucleic acids and proteins can be isolated. Preferred sources include buccal swabs, blood, sperm, epithelial or other tissue, milk, urine, cerebrospinal fluid, sputum, fecal matter, lung aspirates, throat swabs, genital swabs and exudates, rectal swabs, and nasopharyngeal aspirates.

The body fluid sample can be any body fluid which contains cells having the transcript mutation which gives rise to the frameshift mutation and causes or is associated with the diseases. When the disease is a neurodegenerative disease it is preferred that the body fluid sample contains cells of the neuronal system or the products of such cells. When the disease is a neurodegenerative disease, the preferred body fluid is cerebral spinal fluid, which can be obtained after a lumbar puncture (Lannfelt *et al.*, Nature Medicine, 1, pp 829-832, 1995). Another preferred body fluid is blood (including, but not limited to, venous, arterial and cord blood), as it is easily obtained and contains lymphocytes which can be analyzed for the presence of the mutant RNA molecule or encoded protein.

The tissue sample can be any tissue and is preferably one that can be easily obtained, such as skin and nose epithelium.

Preferably, when analyzing the sample for a mutant RNA molecule, a nucleic acid probe is used. The nucleic acid probe is preferably a nucleotide probe having a sequence complementary to part of the mutant RNA molecule encompassing the mutation giving rise to the frameshift mutation.

The probe must be used to detect RNA or DNA reverse transcribed from the RNA, but must not be used to detect genomic DNA as the genomic DNA will not contain the mutation.

The present invention further provides a nucleic acid probe having a sequence complementary to part of the mutant RNA molecule encompassing the mutation leading to the frameshift mutation. The probe is preferably sufficiently complementary to the mutant sequence of the RNA molecule so that under stringent conditions the probe only remains bound to the mutant sequence, and is able to distinguish under stringent conditions the mutant and corresponding wild-type transcripts. "Stringent" conditions are defined herein as RNA:DNA hybridization conditions which may be performed at 65°C using a hybridization buffer equivalent to 50% formamide and 0.1X SSC (see below and Evans et al. PNAS (1994) 9; 6059-6063, 6060). "Stringent" conditions also preferably include stringent washes, as described in Evans et al. (Ibid).

The probe may be of any length but is preferably between 5 and 50 nucleotides long, more preferably between 10 and 30 nucleotides long. For example, the probe may be 5, 10, 15, 20, 25, or 30 nucleotides in length.

In a preferred embodiment the probe comprises a sequence complementary to a GAGA or GAGAX or to a CTCT or CTCTX, having a nucleotide or dinucleotide deletion or insertion, and nucleotide sequences corresponding to the nucleotide sequences flanking the GAGA or CTCT motif in the wild-type RNA molecule. It would be apparent to one skilled in the art that if reverse transcribed DNA complementary to the mutant RNA sequence was being probed for, a probe comprising a sequence complementary to the corresponding GAGA or CTCT motif present in the complementary DNA would have to be used.

Methods of detecting the presence of the mutant RNA molecule include the

reverse transcriptase polymerase chain reaction (RT-PCR) using primers having a sequence complementary to the sequence either side of the mutation which gives rise to the frameshift mutation. Firstly, one primer is used to reverse transcribe the RNA into DNA, and secondly, two primers are used to amplify the DNA, as described hereinbelow.

The primers used in the above RT-PCR based method can vary in size from 20bp to 2-3 kb; for example, 20bp, 50bp, 100bp, 500bp, 1000bp, 1500bp, 2000bp, or 3000bp. The primers can be prepared by a number of standard techniques including cloning the sequences flanking the nucleotide region to be amplified or by synthesizing the primers using phosphoramidite method.

The present invention further provides primers for use in the above defined RT-PCR based methods for the amplification of the nucleotide region containing the mutation.

Preferably, when analyzing the sample for the mutant protein of the present invention an immunological test is employed. The immunological test is preferably based on the use an antibody molecule having specificity for the mutant protein of the present invention and not the wild-type protein.

The present invention thus further provides an antibody molecule having specificity for the mutated protein of the present invention but not for the wild-type protein. Preferably, the antibody is specific for the carboxy terminal end of the mutant protein.

The present invention further provides a method for the diagnosis of a neurodegenerative disease or other age-related diseases, or a method for the diagnosis of a person with a susceptibility for these diseases comprising: i. providing a body fluid or tissue sample from a patient; and ii. analyzing the sample for the presence of an RNA molecule of the neuronal system having a frameshift mutation or a protein encoded thereby, wherein the presence of the mutated RNA molecule is indicative of a neurodegenerative disease.

Preferably, the neurodegenerative disease is AD and Downs' syndrome.

The present invention also relates to methods for preventing and/or treating

the diseases, vectors for preventing and/or treating the diseases and for the production of diagnostic reagents, compositions for preventing and/or treating the diseases, nucleic acid sequences, probes and antibody molecules for use in the present invention and transgenic animals.

Therapies contemplated according to the invention include providing to a cell containing a mutant transcript a ribozyme which is capable of selectively eliminating (i.e., cleaving) the mutant transcript, thus rendering the transcript untranslatable.

The therapies also may include providing to a cell which is thus treated with a ribozyme a corresponding wild-type transcript which is substantially uncleavable by the ribozyme. The wild-type transcript may contain the wild-type sequence corresponding to the mutant RNA sequence, except for the GAGA or CTCT permutation, and encoding the wild-type protein, and also may include third base (in a codon) silent mutations which further differentiate the wild-type RNA from the mutant RNA sequence, and thus further distinguish the sequences with respect to ribozyme recognition and cleavage.

Therapies encompassed by the invention also include providing to cells containing a mutant RNA, an RNA or DNA that is complementary to the mutant RNA and able to form a duplex with the mutant RNA that is untranslatable in the cell. The complementary sequence may be the entire length of the mutant RNA, but is preferably a shorter length, for example, 10, 20, 50 or 100 nucleotides in length. The complementary sequence thus may be administered in the form of an oligonucleotide or may be encoded by an expressible sequence contained in a vector, wherein the vector is administered to the cell.

The invention therefore encompasses a pharmaceutical composition comprising a ribozyme that selectively cleaves a target RNA having a GAGA or CTCT mutation admixed with a pharmaceutically acceptable carrier.

The invention also encompasses a pharmaceutical composition comprising a ribozyme that selectively cleaves a target RNA having a GAGA or CTCT mutation and a wild-type analog of an RNA having a GAGA or CTCT sequence giving rise to a frameshift mutation admixed with a pharmaceutically acceptable carrier.

The invention also encompasses a pharmaceutical composition comprising a wild-type analog of an RNA having a GAGA or CTCT sequence giving rise to a frameshift mutation admixed with a pharmaceutically acceptable carrier.

The invention also encompasses a pharmaceutical composition wherein the wild-type analog of an RNA comprises a nucleotide sequence having third base silent mutations.

The invention also encompasses a pharmaceutical composition comprising a single stranded nucleic acid having a sequence that is complementary to an RNA having one or more GAGA or CTCT mutations giving rise to a frameshift mutation admixed with a pharmaceutically acceptable carrier.

The invention also encompasses a pharmaceutical composition comprising the wild-type analog of a mutant protein in admixture with a pharmaceutically acceptable carrier.

The invention also encompasses a vector comprising an expressible gene encoding a ribozyme that selectively cleaves a target RNA having a GAGA or CTCT sequence.

The invention also encompasses a vector comprising an expressible gene encoding a sequence complementary to an RNA having a GAGA or CTCT mutation giving rise to a frameshift mutation.

The invention also encompasses a host cell containing a vector as described herein.

The invention also encompasses a method of treatment and/or prevention of a disease caused by or associated with an RNA having a GAGA or CTCT mutation giving rise to a frameshift mutation, comprising administering the compositions, vectors, or the host cells described above to a patient suffering from or susceptible to the disease.

The invention also encompasses the use of a vector encoding a ribozyme that selectively cleaves a target RNA having a GAGA or CTCT mutation under the control of a promoter in therapy.

The invention also encompasses the use of a vector encoding a ribozyme

under the control of a promoter in the manufacture of a composition for the treatment of a disease caused by or associated with at least one an RNA having one or more GAGA or CTCT mutations giving rise to a frameshift mutation.

The invention also encompasses the use of a vector encoding the sequence complementary to an RNA having one or more GAGA or CTCT mutations giving rise to a frameshift mutation under the control of a promoter in therapy.

The invention also encompasses the use of more than one of the compositions, the vectors, or the host cells described above in any combination in therapy.

The invention also encompasses the use of more than one of the compositions, the vectors, or the host cells described herein in any combination in the treatment and/or prevention of a disease caused by or associated with at least one an RNA having one or more GAGA or CTCT mutations giving rise to a frameshift mutation.

The present invention further provides an early marker for a neurodegenerative disease. The invention provides a diagnostic kit for diagnosing a disease caused by or associated with at least one RNA molecule having one or more transcript mutations giving rise to a frameshift mutation comprising: i. a nucleic acid probe having a sequence complementary to part of the mutant RNA molecule which encompasses the mutation which leads to the frameshift mutation and packaging materials therefor; and ii. means for detecting the probe bound to the mutant RNA molecule.

The present invention further provides a diagnostic kit for diagnosing a disease caused by or associated with at least one RNA molecule having one or more transcript mutations giving rise to a frameshift mutation comprising: i. primers for use in an RT-PCR reaction, the primers having a sequence complementary to the sequence either side of the mutation which gives rise to the frameshift mutation, packaging materials therefor, and reagents necessary for performing an RT-PCR reaction and amplifying the DNA sequence containing the mutation; and ii. means for detecting the amplified DNA sequence containing the mutation.

The present invention further provides a diagnostic kit for diagnosing a disease caused by or associated with at least one RNA molecule having one or more transcript mutations giving rise to a frameshift mutation comprising: i. an antibody molecule having specificity for the mutant protein of the present invention and not the wild-type protein; and ii. means for detecting the antibody molecule bound to the mutant protein.

The antibody molecule and the means for detecting the bound antibody molecule are as defined above.

In a further embodiment of the present invention the diagnostic kit described above additionally comprising: i. an antibody molecule having specificity for the wild-type protein; and ii. means for detecting the antibody molecule bound to the wild-type protein, as a control for diagnosing a disease caused by or associated with at least one RNA molecule having one or more transcript mutations giving rise to a frameshift mutation.

The present invention further provides an RNA molecule having one or more transcript mutations giving rise to a frameshift mutation which causes or is associated with a disease.

The invention further provides several (one or more) RNA molecules encoding the same amino acid sequence up to the GAGA or CTCT motif, and thereafter encoding different sequences. For example, in a single cell, one RNA molecule encoding a frameshifted protein based on, for example, β-app, may contain a mutation at or within the GAGA motif in exon 9 of the RNA sequence and a second RNA molecule encoding β-app may contain a mutation at or within the GAGA motif in exon 10 of the sequence.

The present invention further provides a mutated protein encoded by the mutated RNA molecule found to be indicative of a disease, the mutant RNA molecule having one or more transcript mutations giving rise to a frameshift mutation. Preferably, the mutant protein contains an antigenic epitope specific for the diseased state, examples of which are provided in Table 9.

In a preferred embodiment of the present invention the mutated RNA

molecule encodes a protein comprising at least part of the sequence designated +1 or +2 in any one of Figures 2 to 19, or an immunologically equivalent fragment thereof.

In a preferred embodiment the mutated protein comprises any one of the following individual sequences: RGRTSSKELA [SEQ ID NO: 1]; HGRLAPARHAS [SEQ ID NO: 2]; YADLREDPDRQ [SEQ ID NO: 3]; RQDHHPGSGAQ [SEQ ID NO: 4]; YADLREDPDRQDHHPGSGAQ [SEQ ID NO: 1400]; GGGAQ [SEQ ID NO: 5], GAPRLPPAQAA [SEQ ID NO: 6]; KTRFQRKGPS [SEQ ID NO: 7]; PGNRSMGHE [SEQ ID NO: 8]; EAEGGSRS [SEQ ID NO: 9]; VGAARDSRAA [SEQ ID NO: 10]; HDYPPGGSV [SEQ ID NO: 11]; SIQKFQV [SEQ ID NO: 12]; VEKPGERGGR [SEQ ID NO: 13]; PLFGRGHKRG [SEQ ID NO: 14]; EDRGDAGWRGH [SEQ ID NO: 15]; QERGASPRAAPREH [SEQ ID NO: 16]; RQPGDVAPGGQHRPVDD [SEQ ID NO: 17]; AGLLAIPEAK [SEQ ID NO: 18]; YVDVYNGGKFS [SEQ ID NO: 19]; AADERRCHLLHMCGRR [SEQ ID NO: 20; QQATEAGQHYQPGSPLHDHSHV [SEQ ID NO: 21]; PQEAAARTNR [SEQ ID NO: 22]; RSWVHPAPPYQMCLG [SEQ ID NO: 23]; and GGSRTHPR [SEQ ID NO: 24], especially when the disease is a neurodegenerative disease such as AD.

In a preferred embodiment, the antibody molecule of the present invention has affinity for the mutant proteins defined above.

The present invention also relates to a method for treating and/or preventing a disease caused by or associated with at least one RNA molecule having one or more transcript mutations giving rise to a frameshift mutation. The finding of mutations in RNA molecules which lead to the production of mutant proteins, and which are indicative of a disease, has led to a number of ways of treating and/or preventing the disease.

The present invention further provides a method for identifying diseases caused by or associated with at least one RNA molecule having one or more transcript mutations giving rise to a frameshift mutation. The method comprises: i. providing the sequence of an RNA molecule suspected of being involved in the pathogenesis of a disease; ii. identifying the sequence of the mutant protein encoded by the RNA sequence 3'-terminal to a frameshift mutation; iii. preparing a probe to the mutant

protein or a fragment thereof; and iv. probing a body fluid or tissue sample from a patient having the disease and a patient not having the disease, in order to find a correlation between the presence of the mutant protein and the diseased state.

Preferably, the probe is an antibody molecule as defined herein. It is further preferred that the antibody molecule has affinity for a protein comprising at least one of the sequences: RGRTSSKELA [SEQ ID NO: 1]; HGRLAPARHAS [SEQ ID NO: 2]; YADLREDPDRQ [SEQ ID NO: 3]; RQDHHPGSGAQ [SEQ ID NO: 4]; YADLREDPDRQDHHPGSGAQ [SEQ ID NO: 1400]; GGGAQ [SEQ ID NO: 5], GAPRLPPAQAA [SEQ ID NO: 6]; KTRFQRKGPS [SEQ ID NO: 7]; PGNRSMGHE [SEQ ID NO: 8]; EAEGGSRS [SEQ ID NO: 9]; VGAARDSRAA [SEQ ID NO: 10]; HDYPPGGSV [SEQ ID NO: 11]; SIQKFQV [SEQ ID NO: 12]; VEKPGERGGR [SEQ ID NO: 13]; PLFGRGHKRG [SEQ ID NO: 14]; EDRGDAGWRGH [SEQ ID NO: 15]; QERGASPRAAPREH [SEQ ID NO: 16]; RQPGDVAPGGQHRPVDD [SEQ ID NO: 17]; AGLLAIPEAK [SEQ ID NO: 18]; YVDVYNGGKFS [SEQ ID NO: 19]; AADERRCHLLHMCGRR [SEQ ID NO: 20; QQATEAGQHYQPGSPLHDHSHV [SEQ ID NO: 21]; PQEAAARTNR [SEQ ID NO: 22]; RSWVHPAPPYQMCLG [SEQ ID NO: 23]; and GGSRTHPR [SEQ ID NO: 24], especially when the disease is a neurodegenerative disease such as AD.

Other features and advantages of the invention will be apparent from the following description of the preferred embodiments thereof, and from the claims.

BRIEF DESCRIPTION OF DRAWINGS

The invention is now illustrated in the appended example with reference to the following drawings:

Figure 1 is a copy of a paraffin section of the frontal cortex of a female Alzheimer patient (70 years old, #83002; Table 2) immunocytochemically incubated with an antibody against a peptide predicted by the + 1 reading frame of βAPP (Figure 20). Dystrophic neurites (arrowheads) and tangles (arrows) are clearly visible in the cortical layer III.

Figure 2 presents the coding nucleotide sequence of the human β amyloid

precursor protein gene transcript [SEQ ID NO: 25], the amino acid sequence of the wild-type protein [SEQ ID NO: 83 and 84], the mutant + 1 frameshift protein [SEQ ID NO: 50-82] and the mutant + 2 frameshift protein [SEQ ID NO: 26-49].

Figure 3 presents the coding nucleotide sequence of the human microtubule-associated protein tau gene transcript [SEQ ID NO: 85], the amino acid sequence of the wild-type protein [SEQ ID NO: 99], the mutant + 1 frameshift protein [SEQ ID NO: 86-98] and the mutant + 2 frameshift protein [SEQ ID NO: 100-112].

Figure 4 presents the coding nucleotide sequence of the human ubiquitin B gene transcript [SEQ ID NO: 113], the amino acid sequence of the wild-type protein [SEQ ID NO: 125 and 126], the mutant + 1 frameshift protein [SEQ ID NO: 114-124] and the mutant + 2 frameshift protein [SEQ ID NO: 127-136].

Figure 5 presents the coding nucleotide sequence of the human apolipoprotein E gene transcript [SEQ ID NO: 137], the amino acid sequence of the wild-type protein [SEQ ID NO: 144-146], the mutant + 1 frameshift protein [SEQ ID NO: 138-143] and the mutant + 2 frameshift protein [SEQ ID NO: 147-152]. Information concerning restriction enzyme sites is also given.

Figure 6 presents the coding nucleotide sequence of the human microtubule-associated protein 2 transcript [SEQ ID NO: 153], the amino acid sequence of the wild-type protein [SEQ ID NO: 154-158], the mutant + 1 frameshift protein [SEQ ID NO: 232-347] and the mutant + 2 frameshift protein [SEQ ID NO: 159-231].

Figure 7 presents the coding nucleotide sequence of the human neurofilament subunit NF-low transcript [SEQ ID NO: 348], the amino acid sequence of the wild-type protein [SEQ ID NO: 466-513], the mutant + 1 frameshift protein [SEQ ID NO: 413-465] and the mutant + 2 frameshift protein [SEQ ID NO: 349-412].

Figure 8 presents the coding nucleotide sequence of the human neurofilament subunit NF-M transcript [SEQ ID NO: 514], the amino acid sequence of the wild-type protein [SEQ ID NO: 515-574], the mutant + 1 frameshift protein [SEQ ID NO: 629-695] and the mutant + 2 frameshift protein [SEQ ID NO: 575-628].

Figure 9 presents the coding nucleotide sequence of the human neurofilament subunit NF-H gene transcript [SEQ ID NO: 696], the amino acid sequence of the

wild-type protein [SEQ ID NO: 697-698], the mutant + 1 frameshift protein [SEQ ID NO: 708-710] and the mutant + 2 frameshift protein [SEQ ID NO: 699-707].

Figure 10 presents the coding mRNA nucleotide sequence [SEQ ID NO: 711] and amino acid sequence of presenilin I expressed in the wildtype [SEQ ID NO: 712], +1 [SEQ ID NO: 733-753], and +2 [SEQ ID NO: 713-732] reading frames.

Figure 11 presents the coding mRNA nucleotide sequence [SEQ ID NO: 754]and amino acid sequence of presenilin II expressed in the wildtype [SEQ ID NO: 776-787], +1 [SEQ ID NO: 755-775], and +2 [SEQ ID NO: 788-814] reading frames.

Figure 12 presents the coding mRNA nucleotide sequence [SEQ ID NO: 815] and amino acid sequence of Big Tau expressed in the wildtype [SEQ ID NO: 816-818], +1 [SEQ ID NO: 824-834] and +2 [SEQ ID NO: 819-823] reading frames.

Figure 13 presents the coding mRNA nucleotide sequence [SEQ ID NO: 835] and amino acid sequence of GFAP expressed in the wildtype [SEQ ID NO: 836-852], +1 [SEQ ID NO: 883-914], and +2 [SEQ ID NO: 853-882] reading frames.

Figure 14 presents the coding mRNA nucleotide sequence [SEQ ID NO: 915] and amino acid sequence of P53 expressed in the wildtype [SEQ ID NO: 940-949], +1 [SEQ ID NO: 916-939] and +2 [SEQ ID NO: 950-965] reading frames.

Figure 15 presents the coding mRNA nucleotide sequence [SEQ ID NO: 966] and amino acid sequence of BCL2 expressed in the wildtype [SEQ ID NO: 967-1015], +1 [SEQ ID NO: 1075-1126] and +2 [SEQ ID NO: 1016-1074] reading frames.

Figure 16 presents the coding mRNA nucleotide sequence [SEQ ID NO: 1127] and amino acid sequence of Semaphorin III expressed in the wildtype [SEQ ID NO: 1128-1131], +1 [SEQ ID NO: 1162-1212] and +2 [SEQ ID NO: 1132-1161] reading frames.

Figure 17 presents the coding mRNA nucleotide sequence [SEQ ID NO: 1213] and amino acid sequence of HUPF expressed in the wildtype [SEQ ID NO: 1241-1244], +1 [SEQ ID NO: 1214-1240] and +2 [SEQ ID NO: 1245-1281]reading frames.

Figure 18 presents the coding mRNA nucleotide sequence [SEQ ID NO:

1282] and amino acid sequence of HMG expressed in the wildtype [SEQ ID NO: 1297-1299], +1 [SEQ ID NO: 1289-1296] and +2 [SEQ ID NO: 1283-1288] reading frames.

Figure 19 presents the coding mRNA nucleotide sequence [SEQ ID NO: 1300] and amino acid sequence of NSP-A expressed in the wildtype [SEQ ID NO: 1374-1387], +1[SEQ ID NO: 1339-1373], and +2 reading frames [SEQ ID NO: 1301-1338].

Figure 20 presents the partial mRNA nucleotide sequence and amino acid sequence of two human neuronal proteins (β amyloid precursor protein (exons 9 and 10) and Ubiquitin B (exon 2)) expressed in the wildtype and +1 reading frame.

Figure 21: Two examples of novel restriction sites generated by dinucleotide deletion in transcripts of β amyloid precursor protein and ubiquitin B (wild-type nucleotide sequences, [SEQ ID NO: 25 and 113];β amyloid precursor protein deletion sequences [SEQ ID NO: 1396 and 1397]; ubiquitin deletion sequences [SEQ ID NO: 1398-1399].

DESCRIPTION

The invention is illustrated by the following nonlimiting examples wherein the following materials and methods are employed. The entire disclosure of each of the literature references cited hereinafter are incorporated by reference herein.

The present invention is based on the discovery that frameshift mutations occur in a single RNA molecule or number of RNA molecules whose product or products are mutant proteins that are associated with, and indicative of, a disease state. The invention is based on the recognition that the presence of a frameshift mutation results in a new coding sequence for the cell containing the frameshift mutation, and thus a new polypeptide (herein termed a mutant protein) which may be correlated with and thus be indicative of a disease.

According to the present invention, diagnosis and/or identification of a disease caused by or associated with at least one RNA molecule having one or more transcript mutations which give rise to a frameshift mutation is accomplished as

described herein.

According to the present invention, methods for preventing and/or treating the diseases, vectors for preventing and/or treating the diseases and for the production of diagnostic reagents, compositions for preventing and/or treating the diseases, nucleic acid sequences, probes and antibody molecules for use in the present invention and transgenic animals are accomplished as described herein.

According to the present invention, methods for detecting errors in transcriptional mechanisms are accomplished as described herein. The correction of the mutations found in the mutant RNA molecules of the present invention is therefore a valuable method for combatting diseases.

Methods and reagents for disease diagnosis and treatment are described in more detail hereinbelow.

Diagnosis of Diseases According to the Invention

The invention relates to methods for diagnosing diseases caused by or associated with at least one RNA molecule having one or more transcript mutations which give rise to a frameshift mutation. Such diseases include but are not limited to cancers, Diabetes mellitus type II and neurodegenerative diseases such as Parkinson's Disease (PD), Alzheimer's Disease (AD), frontal lobe dementia (Pick's Disease), progressive supranuclear palsy (PSP) and other diseases with abundant tau-positive filamentous lesions such as Corticobasal degeneration, Dementia pugilistica, Dementia with tangles only, Dementia with tangles and calcification, Frontotemporal dementias with Parkinsonism linked to chromosome 17, Gertsmann-Strässler-Scheinker disease with tangles, Myotonic dystrophy, Niemann-Pick disease type C, Parkinsonism-dementia complex of Guam, Postencephalitic Parkinsonism, Subacute sclerosing panencephalitis, amyotrophic lateral sclerosis, Huntington's Disease, dementia with Lewy bodies, multisystem atrophy, other inclusion body diseases associated with ubiquitin such as Alexander's disease, Alcoholic liver disease, lichen amyloidosis, and during aging Marinesco bodies and Hyaline inclusions, multiple sclerosis, Downs' syndrome, and other degenerative diseases such as cardiovascular diseases rheumatoid

arthritis, and Diabetes mellitus type II.

Somatic mutations can result in a different gene function and have been implicated in diseases associated with aging, such as certain cancers. However, it has generally been assumed that non-proliferating cells do not undergo important changes at the genomic level. For example, it was assumed previously that genomic changes are mainly related to cell proliferation (Smith, Mutation Research, 277, pp 139-142, 1992) which for non-proliferating cells such as most neurons ends during early postnatal life (Rakic, Science, 277, pp 1054-1056, 1985). However, Evans et al., 1994, Proc. Nat. Acad. Sci. 91:6059, suggested that somatic mutations do occur in genes of the neuronal system, i.e., in post-mitotic neurons. The di/di Brattleboro rat, which suffers from severe diabetes insipidus due to the absence of the antidiuretic hormone vasopressin (VP), was the subject of the Evans et al. paper. It had previously been established that the VP hormone was absent in the Brattleboro rat due to a deletion of a single G residue in the second exon of the VP gene, resulting in a mutant VP precursor with an altered C-terminal amino acid sequence. It had also been observed that a small number of neurons in the di/di rat exhibited a heterozygous +/di phenotype and expressed an apparently normal VP gene product. In studying the molecular biology of the di/di rat, Evans et al. identified sequence alterations that restored the reading frame of the mutant VP precursor mRNA, which were based on a di-nucleotide deletion in a GAGAG motif. They correlated the presence of small amounts of normal VP gene product in single magnocellular neurons with a reversion of the mutant gene stemming from a frameshift mutation. Evans et al. concluded that, because +1 frameshift mutations are present in VP transcripts of both wild-type rats and di/di rats, the events leading to these mutations are not caused by the diseased state of the di/di rat per se. Thus, Evans et al. did not correlate a mutational GAGAG hotspot with a disease state, or predeliction to a disease. Furthermore, there is no suggestion in the prior art that transcript mutations are occurring and that such transcript mutations are caused by or associated with a disease. As the mutations have previously been considered to occur in DNA, methods of detection have been unreliable as there will be no mutation in the genomic DNA and the probing of

genomic DNA will give a false indication of the absence of the mutation.

In the present invention, the observations of Evans et al., as to reversions in the wild-type reading frame at GAGAG hotspots in VP transcripts within single neurons of the di/di rat leading to wild-type-like VP gene products, is extended and developed. According to the present invention, a human disease which is caused by or associated with at least one RNA molecule having one or more transcript mutations occurring at a mutational hotspot and which give rise to a frameshift mutation is identified and/or diagnosed. The nucleotide sequence of an RNA molecule suspected of being involved in the pathogenesis of a disease is provided, e.g., from published gene sequences or from cloning and sequencing of a suspect RNA molecule. The amino acid sequence encoded by the RNA molecule is then predicted, as are amino acid sequences of encoded mutant proteins. Mutant protein sequences are predicted in +1 and +2 reading frames following a hypothesized frameshift mutation. The location of the frameshift mutation may be hypothesized with respect to certain nucleotide sequence motifs which are suspected of causing frameshift mutations, examples of such motifs present in the RNA molecule including but not limited to those comprising GAGA, for example, GAGAC, GAGAG, GAGAT, and GAGAA, or those comprising CTCT, for example CTCTC, CTCTG, CTCTA and CTCTT.

A probe is then prepared that is specific for the mutant protein or an immunogenic fragment thereof (such probes being described hereinabove for detection of proteins or protein fragments). Depending on where the mutation that leads to the frameshift occurs, part of the mutant protein will have the same sequence as the wild-type protein and part of the protein will have the sequence of the mutant protein. Furthermore, depending on where the mutation occurs the mutant protein will terminate when the nucleotide sequence codes for a stop codon (indicated as * in the Figures). Thus, different mutant proteins will be produced depending on where the mutation occurs.

Alzheimer's Disease (AD) is a representative disease diagnosable and treatable according to the invention. AD is a neurodegenerative disease characterised by idiopathic progressive dementia and is the fourth highest major cause of death in

developed countries. It affects 5 to 11% of the population over the age of 65 and as much as 47% of the population over the age of 85. At present there are an estimated 4 million patients suffering from AD in the U.S.A. (see Coleman, Neurobiol. of Aging, 15, Suppl. 2, pp 577-578, 1994), and an estimated 20 million Alzheimer's patients worldwide.

The clinical criteria for AD diagnosis have been defined (see Reisberg et al., Am. J. Psych. 12, pp 1136-1139, 1982; McKhann *et al.*, Neurology, 34, pp 939-944, 1984). The early symptoms of AD vary but generally include depression, paranoia and anxiety. There is also a slow degeneration of intellectual function and memory. In particular, cognitive dysfunction and specific disturbances of speech (aphasia), motor activity (apraxia), and recognition of perception (agnosia) can occur.

There is not yet general consensus in a test for *ante mortem* diagnosis for AD due to the lack of knowledge of the pathogenic mechanisms involved in AD. Diagnosis of AD is made by examination of brain tissue. Such diagnosis is usually carried out on individuals *post mortem*. The diagnosis is based on the presence of a large number of intraneuronal neurofibrillary tangles and of neuritic plaques in the brain tissue, in particular in the neocortex and hippocampus. In order to identify the various types of plaques (e.g. neuritic plaques), neuropil threads and neurofibrillary tangles, staining and microscopic examination of several brain tissue sections is necessary. Neuritic plaques are believed to be composed of degenerating axons (e.g., neuropil threads), nerve terminals and possibly astrocytic and microglial elements. It is also often found that neuritic plaques have an amyloid protein core. The neurofibrillary tangles comprise normal and paired helical filaments and are believed to consist of several proteins.

There are two major types of AD, late onset (>65 years) and early onset (<65 years). Approximately 85% of all AD cases are late onset and only 15% are early onset. Of the latter group 0.3% consists of the hereditary type of AD linked to chromosome 21, 2% of the cases are considered to be linked to chromosome 14, and chromosome 1 has been established for juvenile onset (<0.1%), as discussed below. Sporadic cases are the most prominent group (40%) in early onset AD.

In the most common late onset group, 40% of cases are considered to be familial, meaning that Alzheimer was observed in first degree relatives. Of this familial form only 10% is autosomal dominant. The remaining late onset cases (60%) are non-familial or "sporadic" cases (see Table 1). For these cases relatively little is known and previously no data was available which suggested a possible cause of AD.

At present, it is unclear whether the formation of neuritic plaques and/or neurofibrillary tangles is directly responsible for causing AD. The formation of neuritic plaques, neuropil threads and/or neurofibrillary tangles may be a consequence of a more fundamental cellular or biochemical change.

Diagnostic methods of the invention will include the detection of nucleic acid sequences, preferably via procedures which involve formation of a nucleic acid duplex between two nucleic acid strands, i.e., a nucleic acid probe and a complementary sequence in the mutant RNA or the DNA reverse transcribed from the mutant RNA isolated from a biological sample, or detection of a protein, preferably a mutant or hybrid wild-type/nonsense protein, as defined herein.

1. Preparation and Detection of RNA for Genetic Screening.

Typically, RNA is prepared from the biological sample by DNA extraction procedures well-known in the art (see, e.g., Sambrook et al., 1990, A Laboratory Manual for Cloning, Cold Spring Harbor Press, CSH, NY), and may be further purified if desired, e.g., by electro-elution, prior to analysis.

Methods of detecting a mutant RNA molecule from a biological sample include, but are not limited to the following: (1) reverse transcriptase polymerase chain reaction (RT-PCR) followed by sizing gel electrophoresis or hybridization with an allele-specific (or sequence-specific) probe; (2) hybridization of the eluted RNA with a nucleic acid probe that is complementary to the mutated RNA; (3) the ARMS test, in which one primer has a complementary sequence encompassing the mutation which gives rise to the frameshift mutation, and amplification only occurs if the mutated sequence is present; (4) nucleotide sequencing; (5) RNA amplification via RT-PCR and T7 polymerase; and (6) by a dinucleotide deletion in the RNA after RT-PCR a

cDNA can be generated with novel restriction sites (Figure 21).

A nucleic acid probe useful according to the invention is preferably sufficiently complementary to the mutant sequence of the RNA molecule so that under stringent conditions the probe only remains bound to the mutant sequence (see Evans *et al.*, Proc. Natl. Acad. Sci. USA, 91:6059-6063 (1994). The probe is preferably labelled using any of the standard techniques known to those skilled in the art, such as radioactively using ^{32}P or any other standard isotopes, or using non-radioactive methods including biotin or DIG labelling. The labelled probe can then be easily detected by methods well known to those skilled in the art.

An alternative method for detecting the presence of the mutant RNA molecule is via the reverse transcriptase polymerase chain reaction (RT-PCR). Primers having a sequence complementary to the sequence either side of the mutation which gives rise to the frameshift mutation are used to reverse transcribe the RNA and amplify the reverse transcribed DNA containing the mutation. The mutation in the amplified fragment can then be detected using the probe described above using standard techniques or by sequencing the amplified fragment. The advantages of using the RT-PCR reaction is that less starting material is required and the PCR methods allow quantitative as well as qualitative determinations to be made. Quantitative determinations allow the number of copies of a mutated RNA molecule present in a particular sample to be estimated, and given this information the severity of the diseased state can be estimated.

Another alternative method for detecting the presence of the mutant RNA molecule is one in which one primer has a complementary sequence encompassing the mutation which gives rise to the frameshift mutation. Amplification will therefore only occur if the mutated sequence is present. Newton et al., Nucl. Acids. Res. 17:2503, 1989. The method has previously been used in detecting mutations in the gene responsible for cystic fibrosis, and one skilled in the art could easily perform this test for the detection of the mutant RNA or the reverse transcribed DNA corresponding to the mutant RNA of the present invention.

An example of analysis method (1) follows. The RNA is reverse transcribed

and the DNA then amplified, e.g., using PCR, prior to analysis. Specific conditions for any one PCR, i.e. a PCR targeting a particular sequence, or for any one multiplex PCR, i.e. a PCR targeting a particular set of sequences, may vary but will be known to a person of ordinary skill in the art.

Amplification of a mutated or wild-type reverse transcribed DNA sequence can be accomplished directly from an aliquot of the prepared DNA as follows.

25 µl of DNA is aliquotted into a reaction tube containing 25 µl H_20, 50 µl master mix (see below), 0.5 µl Amplitaq (Perkin Elmer Cetus, Norwalk, CT) and 0.5 µl UNG (Perkin Elmer Cetus, Norwalk, CT). A 50 µl master mix comprises 20 mM Tris HCl, pH 8.3, 100 mM KCl, 5 mM $MgCl_2$, 0.02 µmoles each of dATP, dGTP, dCTP, 0.04 µmoles of dUTP, 20 pmoles of each primer (Perkin Elmer Cetus, Norwalk, CT), and 25 µg gelatin.

A fragment characteristic of the selected amplification sequence can then be visualized under ultraviolet light after ethidium bromide staining a 13% polyacrylamide gel in which an aliquot of the amplification has been electrophoresed. Alternatively, hybridization with allele-specific probes can identify the presence of amplified product from either the normal and/or mutant alleles.

2. Preparation and Detection of Protein for Genetic Screening.

Where the biological molecule to be analyzed is a protein, it may be desirable to release the nucleic acid from biological sample cells prior to protein elution, or to remove nucleic acid from the sample eluate prior to protein analysis. Thus, the sample or eluate may first be treated to release or remove the nucleic acid by mechanical disruption (such as freeze/thaw, abrasion, sonication), physical/chemical disruption, such as treatment with detergents (e.g., Triton, Tween, or sodium dodecylsulfate), osmotic shock, heat, enzymatic lysis (lysozyme, proteinase K, pepsin, etc.), or nuclease treatment, all according to conventional methods well known in the art.

Where a biological sample includes a mutant protein, the presence or absence of which is indicative of a genetic disease, the protein may be detected using

conventional detection assays, e.g., using protein-specific probes such as an antibody probe. Similarly, where a genetic disease correlates with the presence or absence of an amino acid or sequence of amino acids, these amino acids may be detected using conventional means, e.g., an antibody which is specific for the native or mutant sequence (see Table 9 for examples of amino acid sequences present in mutant proteins).

Any of the antibody reagents useful in the method of the present invention may comprise whole antibodies, antibody fragments, polyfunctional antibody aggregates, or in general any substance comprising one or more specific binding sites from an antibody. The antibody fragments may be fragments such as Fv, Fab and $F(ab')_2$ fragments or any derivatives thereof, such as a single chain Fv fragments. The antibodies or antibody fragments may be non-recombinant, recombinant or humanized. The antibody may be of any immunoglobulin isotype, e.g., IgG, IgM, and so forth. In addition, aggregates, polymers, derivatives and conjugates of immunoglobulins or their fragments can be used where appropriate.

The immunoglobulin source for an antibody reagent can be obtained in any manner such as by preparation of a conventional polyclonal antiserum or by preparation of a monoclonal or a chimeric antibody. Antiserum can be obtained by well-established techniques involving immunization of an animal, such as a mouse, rabbit, guinea pig or goat, with an appropriate immunogen.

Preparation of Antibodies

1. Polyclonal antibodies.

The peptide or polypeptide may be conjugated to a conventional carrier (e.g. thyroglobulin) in order to increases its immunogenicity, and antisera to the peptide-carrier conjugate is raised in rabbits. Coupling of a peptide to a carrier protein and immunizations are performed as described (Dymecki, S.M., et al., J. Biol. Chem 267:4815-4823, 1992). Rabbit antibodies against this peptide are raised and the sera titered against peptide antigen by ELISA or alternatively by dot or spot blotting (Boersma and Van Leeuwen, 1994, Jour. Neurosci. Methods 51:317. At the same

time, the antisera may be used in tissue sections. The sera is shown to react strongly with the appropriate peptides by ELISA, following the procedures of Green et al., Cell, 28, 477-487 (1982). The sera exhibiting the highest titer is used in subsequent experiments.

2. Monoclonal antibodies.

Techniques for preparing monoclonal antibodies are well known, and monoclonal antibodies of this invention may be prepared using a synthetic peptide, preferably bound to a carrier, as described by Arnheiter et al., Nature, 294, 278-280 (1981).

Monoclonal antibodies are typically obtained from hybridoma tissue cultures or from ascites fluid obtained from animals into which the hybridoma tissue was introduced. Nevertheless, monoclonal antibodies may be described as being "raised to" or "induced by" the synthetic peptides or their conjugates.

Particularly preferred immunological tests rely on the use of either monoclonal or polyclonal antibodies and include enzyme linked immunoassays (ELISA), immunoblotting, immunoprecipitation and radioimmunoassays. See Voller, A., Diagnostic Horizons 2:1-7, 1978, Microbiological Associates Quarterly Publication, Walkersville, MD; Voller, A. et al., J. Clin. Pathol. 31:507-520 (1978); U.S. Reissue Pat. No. 31,006; UK Patent 2,019,408; Butler, J.E., Meth. Enzymol. 73:482-523 (1981); Maggio, E. (ed.), Enzyme Immunoassay, CRC Press, Boca Raton, FL, 1980) or radioimmunoassays (RIA) (Weintraub, B., Principles of radioimmunoassays, Seventh Training Course on Radioligand Assay Techniques, The Endocrine Society, March 1986, pp. 1-5, 46-49 and 68-78). For analyzing tissues for the presence of the mutant protein of the present invention, immunohistochemistry techniques are preferably used. It will be apparent to one skilled in the art that the antibody molecule will have to labelled to facilitate easy detection of mutant protein. Techniques for labelling antibody molecules are well known to those skilled in the art (see Harlour and Lane, Antibodies, Cold Spring Harbour Laboratory, pp 1-726, 1989).

Alternatively, sandwich hybridization techniques may be used, e.g., an antibody specific for a given protein. In addition, an antibody specific for a haptenic

group conjugated to the binding protein can be used. Another sandwich detection system useful for detection is the avidin or streptavidin system, where a protein specific for the detectable protein has been modified by addition of biotin. In yet another embodiment, the antibody may be replaced with a non-immunoglobulin protein which has the property of binding to an immunoglobulin molecule, for example Staphylococcal protein A or Streptococcal protein G, which are well-known in the art. The protein may either itself be detectable-labeled or may be detected indirectly by a detectable labeled secondary binding protein, for example, a second antibody specific for the first antibody. Thus, if a rabbit-anti-hybrid wild-type/nonsense protein antibody serves as the first binding protein, a labeled goat-anti-rabbit immunoglobulin antibody would be a second binding protein.

In another embodiment, the signal generated by the presence of the hybrid wild-type/nonsense protein is amplified by reaction with a specific antibody for that fusion protein (e.g., an anti-β-galactosidase antibody) which is detectably labeled. One of ordinary skill in the art can devise without undue experimentation a number of such possible first and second binding protein systems using conventional methods well-known in the art.

Alternatively, other techniques can be used to detect the mutant proteins, including chromatographic methods such as SDS PAGE, isoelectric focusing, Western blotting, HPLC and capillary electrophoresis.

Identification of Diseases According to the Invention

The invention provides methods for identifying diseases caused by or associated with at least one RNA molecule having one or more transcript mutations which give rise to a frameshift mutation.

Diseases are identified according to the invention as follows. The nucleotide sequence of an RNA molecule suspected of being involved in the pathogenesis of a disease is provided, e.g., from published gene sequences or from cloning and sequencing of a suspect gene. The amino acid sequence encoded by the RNA is then predicted, as are amino acid sequences of encoded mutant proteins. Mutant protein

sequences are predicted in +1 and +2 reading frames following a hypothesized frameshift mutation. The location of the frameshift mutation may be hypothesized with respect to certain nucleotide sequence motifs in the RNA molecule, examples of such motifs including, but not limited to, GAGA, for example, GAGAC, GAGAG, GAGAT, and GAGAA, or CTCT, for example CTCTG, CTCTC, CTCTA and CTCTT.

A probe is then prepared that is specific for the mutant protein or an immunogenic fragment thereof (such probes being described hereinabove for detection of proteins or protein fragments). Depending on where the mutation that leads to the frameshift occurs, part of the mutant protein will have the same sequence as the wild-type protein and part of the protein will have the sequence of the mutant protein. Furthermore, depending on where the mutation occurs the mutant protein will terminate when the nucleotide sequence codes for a stop codon (indicated as * in the Figures). Thus, different mutant proteins will be produced depending on where the mutation occurs.

The simplest method of probing for the presence of a particular mutant protein is to make an antibody to that protein or an immunogenic portion thereof. An immunogenic fragment may be synthesized corresponding to the C-terminus of the predicted mutant proteins because even if the mutation occurred at another position in the sequence, the probability that the derived mutant protein contains the peptide sequence is increased. For example, in the β-App encoding RNAs, two different transcript modifications have occurred (i.e., at two different GAGA motifs) which result in two frameshifted proteins having identical C-terminal sequences. Furthermore, the C-terminal region of a protein is more likely to form an epitope than other regions of the protein.

Once a probe is made, a biological sample from a patient having the disease and a biological sample from a patient not having the disease is probed for the presence or absence of the mutant protein, also as described above. Alternatively, several probes may be prepared and the combination of probes used to probe the tissue sample. The presence of the mutant protein in a biological sample from a patient

having the disease and the absence of said mutant protein in a biological sample from a patient not having the disease indicates that the mutant protein is a marker for the disease or susceptibility to the disease.

Treatment of Diseases According to the Invention

The invention also relates to methods for preventing and/or treating diseases, vectors for preventing and/or treating the diseases, and compositions such as nucleic acid sequences and proteins for preventing and/or treating the diseases, which methods and compositions are useful in gene and protein therapies.

The invention includes methods of treatment and/or prevention of a disease caused by or associated with an RNA having a mutation in GAGA or CTCT giving rise to a frameshift mutation in which a ribozyme, a wild-type RNA, or both, an RNA or DNA that is complementary to the mutant RNA and capable of forming a hybrid with the mutant RNA, or a vector comprising a sequence encoding any of these sequences, or the wild-type form of a mutant protein, is administered to a patient suffering from or susceptible to the disease.

Preferred diseases which are treated according to the invention include but are not limited to cancer or a neurodegenerative disease, especially AD, the preferred mutant RNAs of the present invention are those encoding the β amyloid precursor protein, the Tau protein, ubiquitin B, apolipoprotein-E_4 (Apo-E_4), micro-tubule associated protein II (MAP 2), the neurofilament proteins (L, M, H), presenilin I, presenilin II, Big Tau, GFAP, P53, BCL2, semaphorin III, HUPF, HMG and NSP-A, having a deletion, insertion or other modification in the RNA leading to a frameshift mutation.

Ribozymes useful in treatment according to the present invention are preferably hammerhead ribozymes.

A pharmaceutical composition according to the invention will include a therapeutically effective amount of a ribozyme, the wild-type analog of the mutant RNA, or both, or a DNA or RNA that is complementary to the mutant RNA and capable of forming a hybrid untranslatable sequence in vivo, in admixture with a

carrier. A therapeutically effective amount is considered that amount which, when administered to a patient, provides a therapeutic benefit to the patient. Such amounts will generally be in the range of 10 ug-100 mg of therapeutic protein/kg body weight of the patient, preferably 50 ug-10 mg, and most preferably 100 ug-1 mg.

Where vectors are useful according to the invention, the vector may be of linear or circular configuration and may be adapted for episomal or integrated existence in the host cell, as set out in the extensive body of literature known to those skilled in the art. The vectors may be delivered to cells using viral or non-viral delivery systems. The choice of delivery system will depend on whether the substance is to be delivered to a selected central nervous system or neuronal cell type or generally to these cells.

Vectors of the present invention additionally may comprise further control sequences such as enhancers or locus control regions (LCS), in order to lead to more controlled expression of the encoded gene or genes. LCS are described in EP-A-0332667. The inclusion of a locus control region (LC), is particularly preferred as it ensures the DNA is inserted in an open state at the site of integration, thereby allowing expression of the gene or genes contained in the vector. The vectors of the present invention have a wide range of applications in ex vivo and in vivo gene therapy.

Animal Models for Disease Diagnosis and Treatment According to the Invention

The invention also includes stable cell lines and transgenic animals for use as disease models for testing or treatment.

A stable cell line or transgenic animal according to the invention will contain a recombinant gene or genes, also known herein as a transgene, encoding one or more mutations giving rise to a frameshift mutation which causes or is associated with a disease.

The recombinant gene will encode an RNA encoding a mutated protein found to be indicative of a disease. Preferably, the mutant protein will contain an antigenic epitope specific for the diseased state. The recombinant gene may encode a protein

comprising at least part of the sequence designated +1 or +2 in any one of Figures 2 to 9, or an immunologically equivalent fragment thereof.

A cell line containing a transgene encoding a mutant protein, as described herein, is made by introducing the transgene into a selected cell line according to any one of several procedures known in the art for introducing a foreign gene into a cell.

A transgenic animal containing such a transgene includes a rodent, such as a rat or mouse, or other mammals, such as a goat, a cow, etc. and may be made according to procedures well-known in the art.

Transgenic animals are useful according to the invention as disease models for the purposes of research into diseases caused by or associated with at least one gene encoding an RNA containing one or more mutations giving rise to a frameshift mutation, and therapies therefore. By specifically expressing one or more mutant genes, as defined above, the effect of such mutations on the development of the disease can be studied. Furthermore, therapies including gene therapy and various drugs can be tested on the transgenic animals.

Recombinant genes introduced into an animal to make a transgenic animal useful in the invention will include those genes specifically disclosed herein, containing a dinucleotide deletion or insertion relative to the wildtype sequence of the gene, the dinucleotide deletion or insertion being associated with the nucleotide sequence GAGA or CTCT; for example GAGAX or CTCTX, where X is one of G, A, T or C; such as GAGAG, GAGAC, GAGAT and GAGAA or CTCTG, CTCTC, CTCTT and CTCTA. Such transgenes will preferably contain a dinucleotide deletion which is an AG deletion or a GT deletion just adjacent to GAGAG (Figure 20), for example, one or two dinucleotide deletions associated with a GAGA, GAGAG, GAGAC, GAGAT, GAGAA leading to a + 1 or + 2 frameshift mutation respectively. In a similar manner, CTCTX can undergo the same deletion process (ΔCT).

Recombinant transgenes containing such a mutation which are particularly useful in animal models of disease include those associated with neurodegenerative diseases, especially Alzheimer's disease, and include but are not limited to mutant gene sequences disclosed herein encoding mutant β amyloid precursor protein, the

Tau protein, ubiquitin B, apolipoprotein-E_4 (Apo-E_4), microtubule associated protein II (MAP 2), the neurofilament proteins (L, M, H), presenilin I, presenilin II, Big Tau, GFAP, P53, BCL2, semaphorin III, HUPF, HMG and NSP-A (see also Tables 2-8).

It also is contemplated that transgenic animals of the invention may contain transgenes that are controlled via a regulatable and/or a regulated promoter such that the corresponding wildtype protein is expressed during selected stages of development and maturity of the animal and in a selected tissue, and the mutant gene is turned-on when desired. This is particularly desirable where the animal model is of Alzheimer's disease, wherein the mutant protein begins to be expressed later in life of the animal. Thus, if the mutant gene is under the control of a brain-specific inducible promoter, e.g., a neurofilament, aldolase or modified Thy-1 promoter, then onset of the disease may be controlled via expression of the mutant gene.

Transgenic animals according to the invention may be generated to over-express a) human β amyloid precursor protein +1, b) human ubiquitin +1 proteins, c) human neurofilament proteins.

EXAMPLE 1

Described below is an embodiment of the invention involving identification of transcript frameshift mutations in RNA molecules encoding proteins which are present in neuronal tissue, and how such mutations are useful in diagnosis of certain disease states.

The cDNA sequences coding for the human β amyloid precursor protein, Tau, ubiquitin, apolipoprotein E4, MAP 2, the neurofilament subunits low, medium and high, presenilin I, presenilin II, Big Tau, GFAP, P53, BCL2, semaphorin III, HUPF-I, HMG and NSP-A were obtained from various gene sequence databases.

Using the sequence data, the various GAGA or CTCT motifs in the sequences were identified, and deletions were hypothesized and the sequences of the derived mutant proteins predicted, as shown in Figures 2-19. Both the sequences of the +1 and +2 frameshift mutant proteins were predicted.

By examining the sequences of the hypothesized mutant proteins, a peptide corresponding to the C-terminus of the hypothesized mutant proteins was synthesized.

The peptides were synthesized using standard techniques known to those skilled in the art. The peptides having the following sequences were synthesized: RGRTSSKELA [SEQ ID NO: 1]; HGRLAPARHAS [SEQ ID NO: 2]; YADLREDPDRQ [SEQ ID NO: 3]; RQDHHPGSGAQ [SEQ ID NO: 4]; YADLREDPDRQDHHPGSGAQ [SEQ ID NO: 1400]; GGGAQ [SEQ ID NO: 5], GAPRLPPAQAA [SEQ ID NO: 6]; KTRFQRKGPS [SEQ ID NO: 7]; PGNRSMGHE [SEQ ID NO: 8]; EAEGGSRS [SEQ ID NO: 9]; VGAARDSRAA [SEQ ID NO: 10]; HDYPPGGSV [SEQ ID NO: 11]; SIQKFQV [SEQ ID NO: 12]; VEKPGERGGR [SEQ ID NO: 13]; PLFGRGHKRG [SEQ ID NO: 14]; EDRGDAGWRGH [SEQ ID NO: 15]; QERGASPRAAPREH [SEQ ID NO: 16]; RQPGDVAPGGQHRPVDD [SEQ ID NO: 17]; AGLLAIPEAK [SEQ ID NO: 18]; YVDVYNGGKFS [SEQ ID NO: 19]; AADERRCHLLHMCGRR [SEQ ID NO: 20; QQATEAGQHYQPGSPLHDHSHV [SEQ ID NO: 21]; PQEAAARTNR [SEQ ID NO: 22]; RSWVHPAPPYQMCLG [SEQ ID NO: 23]; and GGSRTHPR [SEQ ID NO: 24].

Depending on where the mutation that leads to the frameshift occurs, part of the mutant protein will have the same sequence as the wild-type protein and part of the protein will have the sequence of the mutant protein. Furthermore, depending on where the mutation occurs the mutant protein will terminate when the nucleotide sequence codes for a stop codon (indicated as * in the Figures). Thus different mutant proteins will be produced depending on where the mutation occurs.

It is predicted that mutations will occur at GAGA or CTCT motifs in the cDNA and the sequences of the mutant proteins predicted accordingly.

Peptides were synthesized corresponding to the C-terminus of the predicted mutant proteins because even if the mutation occurred at another position in the sequence the probability that the derived mutant protein contains the peptide sequence is increased. Furthermore, the C-terminus region of a protein is more likely to form an epitope than other regions of the protein.

The uniqueness of the synthesized peptides was confirmed by a gene sequence database search.

Each synthesized peptide was then injected into a rabbit and an antibody

having affinity for the peptide purified. The techniques used to obtain the antibodies are standard techniques known to those skilled in the art.

The antibodies obtained were then tested on autopsy material of frontal cortex, temporal lobe and hippocampus of neuropathologically confirmed AD cases and control non-AD cases. The presence of the antibodies is determined using standard detection methods known to those skilled in the art.

Figure 1 shows the presence of the β amyloid precursor mutant protein (βAPP^{+1}) in the frontal cortex of an Alzheimer patient identified using an antibody against a peptide predicted by the +1 reading frame of β amyloid precursor protein. The antibody used had affinity for a peptide having the following sequence RGRTSSKELA [SEQ ID NO: 1].

The results of other immunoreactive tests performed using the antibodies against the predicted peptides are shown in Tables 2-5.

It can be seen that the presence of the mutant protein can be detected and correlates with the subject having AD. The presence of one or more of the mutant proteins can therefore be seen to be indicative of AD.

Table 6 summarizes the immunoreactivity results within the frontal cortex (area 11), temporal cortex (area 38) and the hippocampus.

Other diseases also may be correlated with the presence of mutant proteins, as defined herein. For example, seven patients with Downs' syndrome were tested according to the invention. Downs' syndrome is trisomy of chromosome 21 which leads to over-expression of β-amyloid precursor protein. We noted that the frontal and temporal cerebral cortex and hippocampus of these patients contained plaques and neurofibrillary tangles, and hypothesized that such over-expression may promote accumulation of transcript mutations in neurons, by frameshift mutations at a GAGAG motif in the over-expressed β-amyloid gene. After immunocytochemical staining of tissue from frontal and temporal cerebral cortex from the Downs' patients with the above-described antibody specific for the amyloid +1 carboxy terminal peptide, immunoreactivity was observed in the neurofibrillary tangles in six of seven patients. Staining was absent in the frontal cortex of the matched controls. Therefore, the

mutant amyloid protein is correlated not only with Alzheimer's disease, but also with other diseases, such as Downs', involving Alzheimer's neuropathology.

It has been found that a number of the mutations occur at GAGA or CTCT motifs. Table 7 shows the presence of the complementary GAGA motifs in various cDNAs of the neuronal system. The motif or, as can be seen from the sequences of Tau and apolipoprotein E4, similar motifs such as GAGAG GAGAC, GAGAA, and GAGAT (in the cDNA) may be associated with the frameshift mutations that lead to or are associated with the disease. The presence of the motif or similar motifs in other RNA molecules may indicate that they are relevant to a disease. It is also possible that other mutations occur that are not associated with such motifs but still lead to frameshift mutations that cause or are associated with a disease.

Table 8 shows the presence of GAGAC motifs in particular RNA molecules of the neuronal system, namely βAPP, Tau and Ubiquitin. This table also indicates, *inter alia*, the chromosomal location of the genes from which the mutant RNA molecules are transcribed and the molecular weight of the longest polypeptide forms encoded by the RNA molecules and the predicted size of the aberrant +1 peptide with its C-terminus against which the antibodies were raised. These peptides were revealed in a Western blot and also identified with a different antibody recognizing an epitope on the unaffected wild-type N-terminus.

EXAMPLE 2

Selection of Antigenic Peptide

Synthetic polypeptides corresponding in sequence to a portion of a mutant protein (whether such peptides are chemically synthesized or are chemically or recombinantly generated fragments of a protein), as described herein, will be useful according to the invention as antigenic peptides for generation of antibodies specific for a mutant protein, provided they possess the following characteristics. The synthetic peptide will include a minimum of 8 and preferably 12-15 amino acid residues, and an optimum length of 20-21 amino acids. The hydrophilicity and antigenic index of the amino acid sequence of the hybrid wild-type/nonsense protein may be determined by Analytical Biotechnology Sciences, Boston, MA, using

computer programming. Potential synthetic peptides useful according to the invention include a stretch of 12-20 amino acids preferably within the carboxy terminal 100-150 amino acids of the hybrid wild-type/nonsense protein.

The amino acid sequence of a selected peptide is searched in a computer database of sequences (e.g., GenBank) to preclude the possibility that at reasonable concentrations, antisera to any one peptide would specifically interact with any protein of a known sequence. Preferred sequences are those which are determined not to have a close homolog (i.e., "close" meaning 80-100% identity).

EXAMPLE 3

Detection of "Mutant" Protein

Another embodiment of this invention relates to an assay for the presence of the "mutant" or mutant protein in a given tissue as indicative of a disease state. Here, an above-described antibody is prepared. The antibody or idiotype-containing polyamide portion thereof is then admixed with candidate tissue and an indicating group. The presence of the naturally occurring amino acid sequence is ascertained by the formation of an immune reaction, as signalled by the indicating group. Candidate tissues include any tissue or cell line or bodily fluid to be tested for the presence of the mutant protein, as described hereinabove.

Expression of a given hybrid wild-type/nonsense protein may be investigated using antiserum prepared in rabbits against a peptide corresponding to a carboxy terminal stretch of amino acids in the hybrid wild-type/nonsense protein as follows.

CMK cells or U3T3 cells are metabolically labeled with ^{35}S-methionine and extracts are immunoprecipitated with antiserum. If the hybrid wild-type/nonsense protein is present in the cells, then a protein species of corresponding molecular weight will be detected in CMK and U3T3 cells. The protein may be localized to the membrane, nucleus or cytoplasm by Western blot analysis of the nuclear, membrane and cytoplasmic fractions, as generally described in Towbin et al., Proc. Natl. Acad. Sci. USA, 76, 4350-4354 (1979). This localization may be confirmed by immunofluorescence analysis to be associated mainly with the plasma membrane.

Metabolic labeling immunoprecipitation, and immunolocalization assays are

performed as described previously (Furth, M.E., et al., Oncogene 1:47-58, 1987; Laemmli, U.K., Nature 227:680-685, 1970; Yarden, Y., et al., EMBO J. 6:3341-3351, 1987; Konopka, J.B., et al., Mol. Cell. Biol. 5:3116-3123, 1985). For immunoblot analysis, total lysates are prepared (using Fruth's lysis buffer) (Fruth, M.E., et al., Oncogene, 1:47-58, 1987). Relative protein concentrations are determined with a colorimetric assay kit (Bio-Rad) with bovine serum albumin as the standard. A protein of lysate containing approximately 0.05 mg of protein was mixed with an equal volume of 2 x SDS sample buffer containing 2 mercaptoethanol, boiled for 5 min., fractioned on 10% polyacrylamide-SDS gels (Konopka, J.B., et al., J. Virol., 51:223-232, 1984) and transferred to immunobilon polyvinyldine difluoride (Millipore Corp., Bedford, MA) filters. Protein blots were treated with specific antipeptide antibodies (see below). Primary binding of the specific antibodies may be detected using anti-IgG second antibodies conjugated to horseradish peroxidase and subsequent chemiluminescence development ECL Western blotting system (Amersham International).

For metabolic labeling, 10^6 cells are labeled with 100 µCi of ^{35}S-methionine in 1 ml of Dulbecco's modified Eagles medium minus methionine (Amersham Corp.) for 16 h. Immunoprecipitation of protein from labeled cells with antipeptide antiserum is performed as described (Dymecki, S.M., et al., J. Biol. Chem 267:4815-4823, 1992). Portions of lysates containing 10^7 cpm of acid-insoluble ^{35}S-methionine were incubated with 1 µg of the antiserum in 0.5 ml of reaction mixture. Immunoprecipitation samples were analyzed by SDS-polylarcylamide gel electrophoresis and autoradiography.

For immunolocalization studies, 10^7 CMK cells are resuspended in 1 ml of sonication buffer (60 mM Tris-HCl, pH 7.5; 6 mM EDTA, 15 mM EGTA, 0.75 M sucrose, 0.03% leupeptin 12 mM phenylmethylsulfonyl fluoride, 30 mM 2-mercaptoethanol). Cells are sonicated 6 times for 10 seconds each and centrifuged at 25,000 xg for 10 min at 4°C. The pellet is dissolved in 1 ml of sonication buffer and centrifuged at 25,000 x g for 10 min at 4°C.

The pellet (nucleus fraction) is resuspended in 1 ml of sonication buffer and

added to an equal volume of 2 x SDS sample buffer. The supernatant obtained above (after the first sonication) is again centrifuged at 100,000 x g for 40 min at 4°C. The supernatant (cytosolic fraction) is removed and added to an equal volume of 2 x concentrated SDS sample buffer. The remaining pellet (membrane fraction) is washed and dissolved in sonication buffer and SDS sample buffer as described above. Protein samples are analyzed by electrophoresis on 10% polyacrylamide gels, according to the Laemmli method (Konopka, J.B., et al., Mol. Cell. Biol. 5:3116-3123, 1985). The proteins are transferred from the gels on a 0.45-μm polyvinylidine difluoride membrane for subsequent immunoblot analysis. Primary binding of antibodies is detected using anti-IgG second antibodies conjugated to horseradish peroxidase.

For immunohistochemical localization of a given protein, if desired, CMK cells or U3T3 are grown on cover slips to approximately 50% confluence and are. washed with PBS (pH 7.4) after removing the medium. The cells are prefixed for 1 min at 37°C in 1% paraformaldehyde containing 0.075% Triton X-100, rinsed with PBS and then fixed for 10 min with 4% paraformaldehyde. After the fixation step, cells are rinsed in PBS, quenched in PBS with 0.1 and finally rinsed again in PBS. For antibody staining, the cells are first blocked with a blocking solution (3% bovine serum albumin in PBS) and incubated for 1 h at 37°C. The cells are then incubated for 1 h at 37°C with antiserum (1:100 dilution or with preimmune rabbit serum (1:100) (see below). After the incubation with the primary antibody, the cells are washed in PBS containing 3% bovine and serum albumin and 0.1% Tween 20 and incubated for 1 hour at 37°C in a fluorescein-conjugated donkey anti-rabbit IgG (Jackson Immunoresearch, Maine), diluted 1:100 in blocking solution.

The coverslips are washed in PBS (pH 8.0), and glycerol is added to each coverslip before mounting on glass slides and sealing with clear nail polish. All glass slides were examined with a Zeiss Axiophot microscope.

EXAMPLE 4

Biological Sample Analysis

The above methods for detection of a given mutant protein or nucleic acid are applicable to analyses involving tissues, cell lines and bodily fluids (e.g. cerebrospinal

liquor or blood, including but not limited to venous, arterial and cord blood) suspected of containing the marker protein.

For example, a sample of CNS tissue suspected of being in a diseased state may be analyzed, it having been previously observed according to the invention that tissue of that particular diseased state contains detectable levels of hybrid wild-type/nonsense proteins relative to healthy tissue.

An aliquot of the suspect sample and a healthy control sample are provided and admixed with an effective amount of an antibody specific for the hybrid wild-type/nonsense protein, as herein described, and an indicating group. The admixture is typically incubated, as is known, for a time sufficient to permit an immune reaction to occur. The incubated admixture is then assayed for the presence of an immune reaction as indicated by the indicating group. The relative levels of the hybrid wild-type/nonsense protein in the suspect sample and the control sample are then compared, allowing for diagnosis of a diseased or healthy state in the suspect sample.

The above types of analyzing for the presence of the hybrid wild-type/nonsense protein may, of course, be performed using analysis for the coding RNA, e.g., via Northern blot or RNA dot blot analysis, both of which are conventional and known in the art.

Disease Treatment According to the Invention

Disease treatment according to the invention contemplates eliminating mutant transcripts. Evidence supporting the presence of transcript mutant RNA molecules is that in homozygous Brattleboro hypothalamus cells, vasopressin cDNAs having the frameshift mutation were observed in 1 in 100 colonies, where as genomic vasopressin DNA having the frameshift mutation was not identified.

In the human hypothalamus no age related increase in the number of vasopressin +1 immunoreactive cells is observed (contrary to that in rat). However, in the fetal period (29-42 weeks of gestation) an enormous increase in the number of +1 immunoreactive cells containing the +1 vasopressin protein is detected. After birth, the number of these cells falls back to just a few. In Downs' syndrome, where βapp

gene expression is very high (5-fold higher than normal), the highest levels of βapp and +1 mutant proteins were also observed, higher than in AD where βapp gene expression is not found to be increased over normal levels. It also has been found that βAPP+1 and UbiB+1 mutant proteins coexist (and are present in tangles and dystrophic neurites) in the same cell. Accordingly, is unlikely that the transient increase is due to a genomic event.

Once an RNA molecule containing a frameshift mutation (i.e., a frameshifted transcript), or a mutant protein is correlated with a disease state, the disease is treatable according to the invention as follows: by administering to a patient in need thereof enzymes which serve to selectively eliminate frameshifted RNA via cleavage, e.g., ribozymes; by administering the wild-type version of the mutant RNA, preferably in substantially uncleavable form, by administering the wild-type version of the hybrid wild-type/nonsense protein; or by administering oligonucleotides or sequences encoding oligonucleotides complementary to a mutant RNA to a cell in order to form a nucleic acid duplex which renders the mutant RNA untranslatable.

A patient in need thereof will include a patient exhibiting symptoms of the disease, even those patients suspected of developing the disease, i.e., who are monitored according to the invention by measuring the a tissue sample, e.g., the cerebrospinal fluid, for the presence of frameshifted peptides (e.g. peptides having an amino acid sequence in the +1 or +2 reading frame).

According to the invention, a ribozyme may be delivered to affected or susceptible cells leading to the cleavage of the mutant RNA and resultant inability of the cell to translate the mutant RNA into mutant protein. The wild-type protein, if not already produced by the cell, may be provided in protein form or via administering the wild-type RNA to the cell along with the ribozyme. The wild-type RNA may be engineered so as to contain a sequence that is distinguishable from the mutant RNA sequence other than simply at the level of the GAGA or CTCT mutation. For example, the mutant RNA may contain third base silent mutations, i.e., which do not change the coding sequence of the RNA, but which render the wild-type RNA less or substantially unsusceptible to cleavage by the ribozyme.

Without being bound by any one theory, it is suggested that decreasing the percentage of mutant RNA and increasing the percentage of the correct protein produced in relation to the hybrid wild-type/nonsense protein will reduce or prevent further progression of the disease, and possibly reverse the diseased state. In addition, it is possible that not every mutant transcript results in a mutant protein that is directly toxic to the neuronal tissue. For example, the mutant protein may be routed to the proteasomal and/or lysosomal system or just secreted (e.g. by the constitutive or regulated pathway)and degraded elsewhere. However, sometimes the mutant protein will be accumulated in the membranes of organelles, for instance in the endoplasmic reticulum, thus disrupting the normal processes of the cellular machinery.

The wild-type version of the mutated RNA encodes the correct protein. When the disease is a neurodegenerative disease, preferred wild-type sequences include the RNAs encoded by the β amyloid precursor protein gene, the Tau gene, the ubiquitin B gene, the apolipoprotein-E_4 gene, the microtubule associated protein II (MAP2) gene, the neurofilament protein genes (L, M and H), the presenilin I and II genes, Big Tau, GFAP, P53, BCL2, Semaphorin III, HUPF-1, HMG and NSP-A. The sequences of these genes are provided herein in the figures. Other preferred wild-type RNAs are encoded by the alpha and beta tubulin genes, the sequences of which are found in Cowan et al., Mol. Cell. Biol., 3, 1738-1745(1983) and Lewis et al., J. Mol. Biol. 182, 11-20(1985), respectively.

When the disease is a non-hereditary cancer, preferred wild-type RNAs are encoded by gene sequences which include but are not limited to the human p53 gene and the BCL-2 gene. Mammalian phosphoprotein p53 has been shown to play an essential role in regulation of cell division and is required for the transition from phase G0 to G1 of the cell cycle. P53 is normally present in very low levels in normal cells and is believed to be a tumor suppressor gene; when present at high levels, p53 has been shown to play a role in transformation and malignancy. P53 gene alleles from normal and malignant tissues have been shown to contain BglII site polymorphism (Buchman et al., 1988, Gene 70:245). The p53 coding region contains several GAGA motifs, e.g., GAGAC at position 1476 of the sequence published in Buchman et al.,

GAGA at position 1498; GAGA at position 1643; and GAGA at position 1713, which motifs present candidate sites for frameshift RNAs according to the invention. A frameshift mutation within a p53 RNA thus may lead to loss of the natural p53 tumor suppressor function. Detection of such a mutation in p53 may be diagnostic of pre-malignancy or malignancy, and treatment as described herein which results in correction of p53 function may restore tumor suppressor function.

In Diabetes mellitus type II, which occurs with increased frequency in aged persons, the islands of Langerhans degenerate possibly as a result of frameshift mutations in various transcripts (e.g., the ubiquitin transcript).

The invention also encompasses methods of combatting diseases caused by at least one an RNA having one or more GA, GT or CT deletions giving rise to a frameshift mutation by targeting the RNA transcript. Thus, it is also contemplated according to the invention that a frameshift mutation within an RNA may be corrected at the level of the frameshifted RNA via cleavage using a ribozyme having specificity for the mutant RNA sequence (see Denman et al., Arch. Of Biochem. Biophysics. 323,71-78,1995), and eliminating the mutant mRNA. The disease associated with the frameshifted RNA is thus treated by administering an appropriate ribozyme, or sequences encoding the ribozyme, to the patient.

Ribozymes of selected specificities may be made as described by Sullenger & Cech, Nature 371: 619-622, 1994), herein incorporated by reference. Ribozymes and sequences encoding ribozymes may be prepared as described by Tuschl et al., Curr. Opin. Struc. Biol. 5:296, 1995 and Wahl et al., Curr. Opin. Struc. Biol. 5:282, 1995.

The invention also encompasses methods of treating diseases caused by or associated with at least one an RNA having one or more GAGA or CTCT mutations giving rise to a frameshift by delivery of complementary oligonucleotides or sequences encoding complementary oligonucleotides to a target cell containing a frameshifted RNA. The oligonucleotides will have a mutant sequence with respect to the region of the mutant RNA containing the GA, GT or CT deletion, and thus may serve to form a hybrid in vivo with the mutant RNA, rendering it untranslatable.

Oligonucleotides with strong target site binding affinity, i.e., with full target site homology are preferred. Also preferred are oligonucleotides between 10 - 30 nucleotides in length and containing a CTCT, CTCTG, CTCTC, CTCTT or CTCTA motif or a GAGA, GAGAG, GAGAC, GAGAT or GAGAA motif.

Disease treatment according to the invention is described below and includes preparation of the administered substance and administration of the substance to a patient suffering from a disease according to the invention. As used herein "substance" refers to any one of the following: a ribozyme, a nucleic acid sequence encoding a ribozyme, a wild-type transcript, an antisense mutant RNA or DNA or a nucleic acid sequence encoding an antisense sequence, a wild type protein encoded by the wild type gene, an antibody specific for the frameshifted (nonsense) protein.

Disease treatment according to the invention may be accomplished as follows. In Example 5, treatment using ribozymes according to the invention is described.

In Example 6, administration of vectors is described. In Example 7, administration of proteins, ribozymes, or nucleic acid vectors using liposomes is described. In Example 8, delivery of these substances across the blood-brain barrier is described. Lastly, in Example 9, methods ofdelivering cells comprising a protein, ribozyme or other nucleic acid, such as a vector bearing a gene expression construct, are described.

EXAMPLE 5

Treatment According to the Invention Using Ribozymes

Preparation and delivery of a ribozyme or nucleic acid sequence encoding a ribozyme which effect or facilitate selective removal of the frameshifted RNA is carried out as follows.

Selective Elimination of mutant transcripts According to the Invention.

The invention thus also encompasses methods of treating diseases caused by the translation of frameshifted mRNA's which are the result of transcriptional infidelity occurring at or adjacent to GAGA or CTCT motifs in the β-APP and ubiquitin B genes. It is believed that accumulation of aberrant proteins encoded by

these messages contributes to the progression of Alzheimer's disease; therefore, elimination of the mutant transcripts is of therapeutic value. It is contemplated according to the invention that the mutant transcripts described herein are rendered untranslatable in a cell *via* ribozyme-mediated cleavage using ribozymes designed and administered as described herein. In addition, it may be advantageous in certain circumstances to replace the ribozyme-cleaved messages with an exogenous transcript encoding the wild-type protein, which transcript is cleavage-resistant and the synthesis of which is therefore not subject to transcriptional errors or post-transcriptional modification such as those that produced the mutant transcripts described herein.

Treatment strategies are described below for selective elimination of the ubiquitin B and β-APP mutant transcripts in cells. The invention, however, also contemplates selective removal of other mutant transcripts, whether disclosed herein or later-discovered, according to the methods described hereinbelow.

Ribozymes of the hammerhead class are the smallest known, and lend themselves both to *in vitro* synthesis and delivery to cells (summarized by Sullivan, *J. Invest. Dermatol. 103*: 85S-98S, 1994; Usman *et al., Curr. Opin. Struct. Biol. 6*: 527-533, 1996). It is required of hammerhead cleavage targets that they comprise the sequence motif UH, wherein H denotes the ribonucleotides A, U, or C, but not G; the sequence is cleaved following the H. The core functional unit of the hammerhead ribozyme is a tripartite structure made up of helix I, which hybridizes to mRNA sequences 3' of the cleavage site, helix II, a 22 ribonucleotide catalytic domain which mediates the cleavage reaction, and helixIII, which hybridizes to sequences 5' of- and including the "U" of the UH cleavage motif (Haseloff and Gerlach, *Nature 334*: 585-591, 1988; Ruffner *et al., Biochemistry 33*: 10695-10702, 1990). Studies have shown that the lengths of helices I and III are proportional to the efficiency with which ribozymes both bind the area surrounding the cleavage site and release themselves from it following cleavage; the former is critical for target recognition, while the latter is important for maintaining kinetics that indicate true catalytic activity, namely, raising the ratio of target molecules inactivated to ribozymes above the 1:1 stoichiometric ratio observed with antisense-RNA-mediated inactivation. Ideally,

helix I is 3 to 5 ribonucleotides in length, relative to 9 to 13 ribonucleotides for helix III (Tabler *et al., Nucleic Acids Res. 22*: 3958-3965, 1994; Hendry and McCall, *Nucleic Acids Res. 24*: 2679-2684, 1996). Other factors, such as stem loop formation in the unbound ribozyme and target mRNA also play a role in reaction kinetics and ribozyme stability, and in order to predict and/or compensate for such interactions, molecular modeling studies and *in vitro* trials of numerous ribozyme designs are often undertaken (Sioud *et al., Nucleic Acids Res. 22*: 5571-5575, 1994; Gavin and Gupta, *J. Biol. Chem. 272*: 1461-1472, 1997).

Mutant ubiquitin B transcripts can be removed from cells *via* the application of hammerhead ribozymes delivered to these cells in liposome vectors. The invention comprises use of these ribozymes to recognize and cleave mutant transcripts at a site 5' to the GAGA or GGT-containing site that is the source of polymerase slippage during transcription, thereby ridding the cells of the frame-shifted portion of the translated products of these defective messages.

The sequence immediately preceding the GAGA motif in the ubiquitin transcript is GCGUCU, which includes the cleavage recognition motif UC. Given the length considerations posed above, the sequence ideally bound by helix I for a particular mutant is GAG; however, in that such a ribozyme would be expected to bind the mutant and wild-type transcripts with equal efficiency, helix I must be lengthened to include four more nucleotides, such that all seven bases will hybridize to the mutant transcript, while sufficient mismatch to destabilize binding to the wild-type sequence will result. This strategy is more efficient in cases in which the mutant transcript has arisen *via* deletion rather than insertion, since in the latter, the effect of the absolute length of helix I on target release becomes a concern; however, delivery of a pool of differentially-designed ribozymes complementary to various mutant sequences that can result from imprecise transcription of the GAGA motif or GGT sequence in the case of ubiquitin and of sufficient mismatch with the wild-type sequence to inhibit efficient binding of a ribozyme to it should eliminate translation of the frame shifted products of a large proportion of defective messages.

Such a strategy is practical in situations in which the cleavage site is 1 to (at

most) 5 bases to the 5' side of the cleavage site; however, longer distances require accordingly longer helix I binding domains which, combined with the need to create 3' mismatches for differentiation between mutant and wild-type transcripts, make such an approach inadequate for dealing with certain mutations. This is true of GAGA-defective transcripts of β-APP. While one GAGA motif is separated from the cleavage site by a single base, the remaining four motifs are between 7 and 20 bases from the nearest 5' cleavage site. In such a case, cleavage of both of the wild-type transcript *via* hammerhead ribozymes may be unavoidable, and its replacement with a cleavage-resistant transcript must be undertaken in concert with removal of mutant transcripts (see Table 10 for possible sequence substitutions resulting in a cleavage-resistant transcript); here, the ribozyme is designed to cleave both types of message at any UH site 5' of the first GAGA motif.

It may be advantageous to replace the β-APP transcript in all cells in which it is needed; therefore, co-delivery of an expression vector bearing a spliced β-APP minigene driven by β-APP promoter sequences may be employed. Numerous studies of this promoter have been undertaken (among them, Lahiri and Robakis, *Brain Res. Mol. Brain Res. 9*: 253-7, 1991; Bourbonniere and Nalbantoglu, *Brain Res. Mol. Brain Res. 19*: 246-250, 1993; Lukiw *et al., Brain Res. Mol. Brain Res. 22*: 121-131, 1994; Lahiri and Nall, *Brain Res. Mol. Brain Res. 32*: 233-40, 1995; Bourbonniere and Nalbantoglu, *Brain Res. Mol. Brain Res. 35*: 304-308, 1996; Quitschke *et al., J. Biol. Chem. 271*: 22231-9, 1996), and it has been demonstrated that 96 base pairs 5' to the transcriptional start site are sufficient for cell-type-specific promoter activity in tissue culture (Quitschke and Goldgaber, *J. Biol. Chem. 267*: 17362-17368, 1992). The 96 base pairs can be fused to a minigene engineered such that alterations are made in the ribozyme recognition site to prevent cleavage of the replacement β-APP transcript and in the GAGA motifs to inhibit slippage of the transcriptional machinery such as produces the mutant transcripts in the first place; these replacements should be performed such that translationally "silent" mutations are introduced in each case. Examples of such changes are shown in Table 10.

EXAMPLE 6

Preparation of Nucleic Acid Vectors

Sequences encoding ribozymes or a wild-type version of a mutant RNA, or an antisense (complementary) mutant oligonucleotide sequence may be cloned into an appropriate vector for expression in a desired mammalian cell. The vector will include a promoter that is expressed in the target cell type, and also may include an enhancer and locus control region, as selected for expression in a given cell type. Examples of vectors useful according to the invention include but are not limited to any vector which results in successful transfer of the coding sequences to the target mammalian cell. A nucleic acid may be transfected for use in the invention using a viral (e.g. adenoviral or retroviral) or non-viral DNA or RNA vector, where non-viral vectors include, but are not limited to, plasmids, linear nucleic acid molecules, artificial chromomosomes and episomal vectors. Expression of heterologous genes has been observed after injection of plasmid DNA into muscle (Wolff J. A. et al., 1990, Science, 247: 1465-1468; Carson D.A. et al., US Patent No. 5,580,859), thyroid (Sykes et al., 1994, Human Gene Ther., 5: 837-844), melanoma (Vile et al., 1993, Cancer Res., 53: 962-967), skin (Hengge et al., 1995, Nature Genet., 10: 161-166), liver (Hickman et al., 1994, Human Gene Therapy, 5: 1477-1483) and after exposure of airway epithelium (Meyer et al., 1995, Gene Therapy, 2: 450-460).

For example, the retroviral gene transfer vector SAX (Kantoff et al., Proc. Nat. Aca. Sci. 83:6563, 1986) may be used to insert a selected coding sequence into a target cell. SAX is a moloney virus based vector with the neoR gene promoted from the retroviral LTR and the human ADA gene promoted from an internal SV40 promoter. Thus, the SAX vector may be engineered by one of skill in the art to contain the coding sequence for a ribozyme, or a wild-type RNA, or a selected antisense sequence, identified as described herein, e.g., by substituting the desired coding region for the hADA coding region in the SAX vector.

Expression vectors are known in the art which encode, or may be engineered to encode, a selected ribozyme. Yuyama et al., Nucl. Acids Res. 22:5060, 1994, describe a multifunctional expression vector encoding several ribozymes. This vector

may be adapted to encoded a ribozyme of a selected specificity by substituting one or both ribozyme sequences in the vector for a selected ribozyme sequence. Zhou et al., Gene 149:33, 1994, and Yamada et al., Virology 205;121, 1994, describe retroviral transduction of ribozyme sequences into T cells. These retroviral vectors may be adapted to encode a selected ribozyme sequence. Liu et al., Gene Therapy 1:32, 1994, and Lee et al., Gene Therapy 2:377, 1995, describe expression vectors which are adaptable for use in expression of any nucleic acid sequence contemplated according to the invention.

Generally, nucleic acid molecules are administered in a manner compatible with the dosage formulation, and in such amount as will be prophylactically and/or therapeutically effective. When the end product (e.g. an antisense RNA molecule or ribozyme) is administered directly, the dosage to be administered is directly proportional to the the amount needed *per* cell and the number of cells to be transfected, with a correction factor for the efficiency of uptake of the molecules. In cases in which a gene must be expressed from the nucleic acid molecules, the strength of the associated transcriptional regulatory seuqences also must be considered in calculating the number of nucleic acid molecules *per* target cell that will result in adequate levels of the encoded product. Suitable dosage ranges are on the order of, where a gene expression construct is administered, 0.5- to 1μg, or 1- 10μg, or optionally 10- 100 µg of nucleic acid in a single dose. It is conceivable that dosages of up to 1mg may be advantageously used. Note that the number of molar equivalents *per* cell vary with the size of the construct, and that absolute amounts of DNA used should be adjusted accordingly to ensure adequate gene copy number when large constructs are injected.

Nucleic acid molecules to be administered according to the invention may, for example, be formulated in a physiologically acceptable diluent such as water, phosphate buffered saline, or saline, and further may include an adjuvant; however, it is contemplated that other formulations may advantageously be employed. Adjuvants such as incomplete Freund's adjuvant, aluminum phosphate, aluminum hydroxide, or alum are materials well known in the art. Administration of a nucleic acid molecule

as described herein may be either localized or systemic. Methods for both localized and systemic administration of a pharmacological composition are well known in the art.

Nucleic acid constructs of use in the invention can be given in a single- or multiple dose. A multiple dose schedule is one in which a primary course of administration can include 1-10 separate doses, followed by other doses given at subsequent time intervals required to maintain and or reinforce the cellular level of the transfected nucleic acid. Such intervals are dependent on the continued need of the recipient for the therapeutic nucleic acid, the ability of a given nucleic acid to self-replicate in a mammalian cell if it does not become integrated into the recipient's genome and the half-life of a non-renewable nucleic acid (e.g. a molecule that will not self-replicate). Preferably, when the medical needs of the recipient mammal dictate that a nucleic acid or a product thereof will be required throughout its lifetime, or at least over an extended period of time, such as a year or more, a nucleic acid may be encoded by sequences of a vector that will self-replicate in the target cells. The efficacy of transfection and subsequent maintenance of the nucleic acid molecules may be assayed either by monitoring the activity of a marker gene, which may additionally be comprised by the transfected construct, or by the direct measurement of either the protein product encoded by the gene of interest or the reduction in the levels of a protein the production of which it is designed to inhibit. The assays can be performed using conventional molecular and biochemical techniques, such as are known to one skilled in the art.

The success of treatment using nucleic acid molecules in the invention may be determined by the assessment of known clinical indicators (e.g., for a neurodegenerative disease, loss of cognitive or motor function). The progression or (if treatment is undertaken prophylactically on an patient believed to be at risk of disease) development of such symptoms in a treated individual is compared to those observed in untreated control subjects; if an improvement in the treated patient's condition is observed relative to that of control subjects, treatment is judged to be effective. The making of such an assessment is well within the knowledge of one of

skill in the art.

EXAMPLE 7

Liposomal Delivery According to the Invention

Substances may be administered according to the invention using any delivery means known in the art. Described below is liposomal delivery. Liposomes which are used to administer the substances described herein, e.g., a ribozyme can be of various types and can have various compositions. The primary restrictions are that the liposomes should not be toxic to the living cells and that they should deliver their contents into the interior of the cells being treated.

The use of pH sensitive liposomes to mediate the cytoplasmic delivery of calcein and FITC dextran has been described (see Straubinger et al., Cell 32:1069-1079, 1983; and Straubinger et al., FEBS Letters 179:148-154, 1985. Other discussions of pH sensitive liposomes can be found in chapter 11 of the book CELL FUSION, edited by A.E. Sowers, entitled "Fusion of Phospholipid Vesicles Induced by Divalent Cations and Protons" by Nejat Duzgunes et al., Plenum Press, N.Y., 1987, 241-267. See also Ellens et al., Biochemistry, 23:1532-1538, 1984, and Bentz et al., Biochemistry 26:2105-2116, 1987.

The liposomes may be of various sizes and may have either one or several membrane layers separating the internal and external compartments. The most important elements in liposome structure are that a sufficient amount of enzyme or nucleic acid be sequestered so that only one or a few liposomes are required to enter each cell for delivery of the substance, and that the liposome be resistant to disruption. Liposome structures include small unilamellar vesicles (SUVs, less than 250 angstroms in diameter), large unilamellar vesicles (LUVs, greater than 500 angstroms in diameter), and multilamellar vesicles (MLs). In the example presented below, although SUVs are used to administer a ribozyme, the methods are applicable to administration of any substance described herein.

SUVs can be isolated from other liposomes and unincorporated enzyme by molecular weight can be isolated from other liposomes and unincorporated enzyme by molecular sieve chromatography, which is precise but time consuming and dilutes the

liposomes, or differential centrifugation, which is rapid but produces a wider range of liposome sizes.

The liposomes may be made from natural and synthetic phospholipids, glycolipids, and other lipids and lipid congeners; cholesterol, cholesterol derivatives and other cholesterol congeners; charged species which impart a net charge to the membrane; reactive species which can react after liposome formation to link additional molecules to the liposome membrane; and other lipid soluble compounds which have chemical or biological activity.

The liposomes useful according to the invention may be prepared, for example, as described in U.S. Patent No. 5,296,231, which describes preparation of liposomes containing a ribozyme, although it should be borne in mind that liposomes useful according to the invention may contain any one of the substances as herein described. Briefly, by combining a phospholipid component with an aqueous component containing the ribozyme (or desired substance) under conditions which will result in vesicle formation. The phospholipid concentration must be sufficient to form lamellar structures, and the aqueous component must be compatible with biological stability of the enzyme. Methods for combining the phospholipids onto glass and then vesicles will form include: drying the phospholipids onto glass and then dispersing them in the aqueous component; injecting phospholipids dissolved in a vaporizing or non-vaporizing organic solvent into the aqueous component which has previously been heated; and dissolving phospholipids in the aqueous phase with detergents and then removing the detergent by dialysis. The concentration of the ribozyme in the aqueous component can be increased by lyophilizing the enzyme onto dried phospholipids and then rehydrating the mixture with a reduced volume of aqueous buffer. SUV's can be produced from the foregoing mixtures either by sonication or by dispersing the mixture through either small bore tubing or through the small orifice of a French Press.

Ribozymes incorporated into liposomes can be administered to living cells internally or topically. Internal administration to animals or humans requires that the liposomes be pyrogen-free and sterile. To eliminate pyrogens, pyrogen-free raw

materials, including all chemicals, enzymes, and water, are used to form the liposomes. Sterilization can be performed by filtration of the liposomes through 0.2 micron filters. For injection, the liposomes are suspended in a sterile, pyrogen-free buffer at a physiologically effective concentration. Topical administration also requires that the liposome preparation be pyrogen-free, and sterility is desirable. In this case, a physiologically effective concentration of liposomes can be suspended in a buffered polymeric glycol gel for even application to the skin. In general, the gel should not include non-ionic detergents which can disrupt liposome membranes. Other vehicles can also be used to topically administer the liposomes.

The concentration of the substance in the final preparation can vary over a wide range, a typical concentration being on the order of 50 ug/ml. In the case of pH sensitive liposomes, lower concentrations of the substance can be used, e.g., on the order of 0.01 to 1.0 ug/ml for liposomes administered to cells internally. In case of topical application, higher liposome concentrations used, e.g., ten or more times higher.

EXAMPLE 8

Administration Across the Blood-Brain Barrier

Where it is desired according to the invention to administer a substance as described herein or its coding sequence, or liposomes containing such substances, to an individual such that the administered material crosses the blood-brain barrier, several methods are known in the art.

For example, a substance to be administered, whether it be protein or nucleic acid or liposome, may be co-administered with a polypeptide, for example a lipophilic polypeptide that increases permeability at the blood-brain barrier. Examples of such polypeptides include but are not limited to bradykinin and receptor mediated permeabilizers, such as A-7 or its conformational analogues, as described in U.S. Patent Nos. 5,112,596 and 5,268,164. The permeabilizing polypeptide allows the co-administered ribozyme, coding sequence or liposome to penetrate the blood-brain barrier and arrive in the cerebrospinal fluid compartment of the brain, where the ribozyme, or coding sequence may then reach and enter a target neuronal cell.

Alternatively, the substance to be administered may be coupled to a steroidal estrogel or androgel to increase binding to steroid receptors and thus access to the brain.

Another exemplary method for administering a substance such as a ribozyme, antibody, nucleic acids, or liposomes containing such molecules, according to the invention includes forming a complex between the substance to be administered and an antibody that is reactive with a transferrin receptor, as described in U.S. Patent No. 5,182,107. The complex may include a cleavable or non-cleavable linker and is administered under conditions whereby binding of the antibody to a transferrin receptor on a brain capillary endothelial cell occurs and the substance is transferred across the blood-brain barrier in active form.

EXAMPLE 9

Ex vivo therapy

It is possible to administer a therapeutic nucleic acid for use not only in *in vivo* therapy (i.e., that in which a nucleic acid is administered directly to a patient for uptake by- and subsequent expression in cells *in situ*) but also in *ex vivo* therapy (i.e., that in which a nucleic acid is administered to cultured or explanted cells *in vitro*, which transfected cells are subsequently transplanted into the clinical patient in order to supply a therapeutic product). Methods of *ex vivo* gene therapy are described in detail herein. By these methods, a plasmid which continues to be maintained in a transformed or transfected cell *after* such a cell has been administered (e.g. *via* transplantation) to a multicellular host, such as a mammal, delivers a gene product to that individual. It is contemplated that a gene of interest, particularly a therapeutic gene, will be expressed by the transplanted cell, thereby providing the recipient organism, particularly a human, with a needed RNA (e.g., an antisense RNA or ribozyme) or protein.

A cell type may be used according to the invention which is amenable to methods of nucleic acid transfection such as are known in the art. Such cells may include cells of an organism of the same species as the recipient organism, or even cells harvested from the recipient organism itself for *ex vivo* nucleic acid transfection prior to re-introduction. Such autologous cell transplants are known in the art. One

common example is that of bone marrow transplantation, in which bone marrow is drawn either from a donor or from a clinical patient (for example, one who is about to receive a cytotoxic treatment, such as high doses of ionizing radiation), and then transplanted into the patient *via* injection, whereupon the cells re-colonize bones and other organs of the hematopoietic system.

a. Cell dosage

The number of transfected cells which are administered to a recipient organism is determined by dividing the absolute amount of therapeutic or other gene product required by the organism by the average amount of such an agent which is produced by a transfected cell. Note that steady-state plasmid copy number varies depending on the strength of its origin of replication as well as factors determined by the host cell environment, the availability of nucleotides and replicative enzyme complexes, as does the level of expression of the gene of interest encompassed by the plasmid, which level likewise is determined by the strength of its associated promoter and the availability of nucleotides and transcription factors in a given host cell background. As a result, the level of expression *per* cell of a given gene of interest must be determined empirically prior to administration of cells to a recipient.

While efficient methods of cell transfection and transplantation are known in the art, they do not ensure that the transfected cell is immortal. In addition, the requirements of the recipient organism for the product encoded by the transgene may change over time. In light of these considerations, it is contemplated that cells may be administered in a single dose or in multiple doses, as needed. A multiple dose schedule is one in which a primary course of administration can include 1-10 separate doses, followed by other doses given at subsequent time intervals required to maintain and or reinforce the cellular level of the transfected nucleic acid. Such intervals are dependent on the continued need of the recipient for the therapeutic gene product. Preferably, when the medical needs of the recipient mammal dictate that a gene product will be required throughout its lifetime, or at least over an extended period of time, such as a year or more, the transfected cells will be replenished on a regular schedule, such as monthly or semi-monthly, unless such cells are able to colonize the

recipient patient in permanent fashion, such as is true in the case of a successful bone-marrow cell transplant.

b. Nucleic acid dosage

Provided a nucleic acid vector capable of replication in the transfected cell is used, the absolute amount of nucleic acid which is transfected into cells prior to transplantation is not critical, since in cells receiving at least one copy of such a vector, the vector will replicate until an equilibrium copy-number is achieved. As a first approximation, an amount of vector equivalent to between 1 and 10 copies thereof *per* cell to be transfected may be used; one of skill in the art may adjust the ratio of plasmid molecules to cells as is necessary to optimize vector uptake. Of particular used in the invention are vectors or transfection techniques which result in the stable integration of the gene of interest into the chromosome of the transfected cell, so as to avoid the need to maintain selection for cells bearing the vector following transplantation into a recipient multicellular organism, such as a human.

c. Administration of autologous or syngeneic cells

A cell type which is commonly transplanted between individuals of a single species (or, even, from an individual to a cell culture system and back to the same individual) is that of hematopoietic stem cells (HSCs), which are found in bone marrow; such cells have the advantage that they are amenable to nucleic acid transfection while in culture, and are, therefore, well suited for use in the invention. Cultures of HSCs are transfected with a minimal plasmid comprising an operator sequence and a gene of interest and the transfected cells administered to a recipient mammal in need of the product of this gene. Transfection of hematopoietic stem cells is described in Mannion-Henderson et al., 1995, Exp. Hematol., 23: 1628; Schiffmann et al., 1995, Blood, 86: 1218; Williams, 1990, Bone Marrow Transplant, 5: 141; Boggs, 1990, Int. J. Cell Cloning, 8: 80; Martensson et al., 1987, Eur. J. Immunol., 17: 1499; Okabe et al., 1992, Eur. J. Immunol., 22: 37-43; and Banerji et al., 1983, Cell, 33: 729. Such methods may advantageously be used according to the present invention. Administration of transfected cells proceeds according to methods established for that of non-transfected cells, as described below.

The transplantation of hematopoietic cells, such as in a bone marrow transplant, is commonly performed in the art by procedures such as those described by Thomas et al. (1975, New England J. Med., 292: 832-843) and modifications thereof. Such a procedure is briefly summarized: In the case of a syngeneic graft or of a patient suffering from an immunological deficiency, no immunosuppressive pre-treatment regiment is required; however, in cases in which a cells of a non-self donor are to be administered to a patient with a responsive immune system, an immunosuppressive drug must be administered, e.g. cyclophosphamide (50 mg/kg body weight on each of four days, with the last does followed 36 hours later by the transplant). Leukemic patients routinely receive a 1000-rad midline dose of total-body irradiation in order to ablate cancerous blood cells; this irradiation also has an immune-suppressive effect. Following pre-treatment, bone marrow cells (which population comprises a small number of pluripotent hematopoietic stem cells, or HSCs), are administered *via* injection, after which point they colonize the hematopoietic system of the recipient host. Success of the graft is measured by monitoring the re-appearance of the numerous adult blood cell types by the immunological and molecular methods which are well known in the art. While as few as 1-10 HSCs are, in theory, able to colonize and repopulate a lethally-irradiated recipient mammal over time, it is advantageous to optimize the rate at which repopulation occurs in a human bone marrow transplant patient; therefore, a transplanted bone marrow sample comprising 10 to 100, or even 100 to 1000 HSCs should be administered in order to be therapeutically effective.

It is contemplated that both lymphoid and parenchymal cells are of use in the invention. Such parenchymal cells include those of the islets of Langerhans, the thyroid, the adrenal cortex, muscles, cartilagenous- or other synovial tissue, the kidneys, epithelial tissues (both external and internal, particularly that of the intestinal lumen, lung, heart, liver, kidney, neurons and synovial cells) and, in particular, the nervous system.

To render the transplanted cells resistant, at least collectively, to immune rejection by the recipient organism, it is contemplated that transplanted cells expressing a high

level of activated NFκB (a high NFκB "set point"), while still subject to destruction by autoimmune host lymphocytes, would enjoy the advantage of robust proliferative capacity in order to multiply at a rate surpassing that of cell killing, thereby providing a long-lived population of therapeutic cells to the recipient organism. Such cells may be transfected with gene expression constructs which result in the production of high levels of activated NFκB, or may be cells obtained from a donor selected for high endogenous NFκB activity, as may be determined in an *in vitro* transcription assay or DNA/protein binding assay by methods well known in the art, using protein extracts drawn from such a donor, which may, itself, be a transgenic mammal.

d. Administration of xenogeneic and allogeneic cells

While transfection and subsequent tranplantation of cells which are obtained from an individual or cell culture system of like species with the recipient organism may be performed, it is equally true that the invention may be practised using cells of another organism (such as a well-characterized eukaryotic microorganism, e.g. yeast, in which appropriate processing of proteins encoded by therapeutic genes is likely and in which useful origins of replication are known). In such a case, certain concerns must be addressed.

First, when a protein is encoded by the gene of interest, the transplanted cells must produce the protein in a form that may is of use to the recipient organism. Post-translational processing (including, but not limited to, cleavage and patterns of glycosylation) must be consistent with proper function in the recipient. In addition, either a protein or an RNA molecule of interest must be made available to the recipient after synthesis, such as by secretion, excretion or exocytosis from the transplanted cell. To address the former, the protein produced by the transfected cells may be qualitatively compared to the native protein produced by an individual of the same species as the recipient organism by biochemical methods well known in the art of protein chemistry. The latter, release of the protein of interest by the cells to be transplanted, may be assayed by isolating protein from culture medium which has been decanted from the transfected cells or from which such cells have been separated (i.e. by centrifugation or filtration), and performing Western analysis using an antibody

directed at the protein of interest. Antibodies against many proteins are commercially available; techniques for the production of antibody molecules are well known in the art.

Second, the cells must be shielded from immune rejection by the recipient organism. It is contemplated that such cells may be transfected with constructs expressing cell-surface markers (e.g. MHC antigens) characteristic of the recipient patient so as to provide them with biochemical camoflage.

In addition, methods for the encapsulation of living cultures of cells for growth either in an artificial growth environment, such as in a fermentor, or in a recipient organism have been developed, and are also of use in the administration of cells transfected according to the invention. Such an encapsulation system renders the cell invisible to immune detection and, in addition, allows for the free exchange of materials (e.g. the gene product of interest, oxygen, nutrients and waste materials) between the transplanted cells and the environment of the host organism.

Methods and devices for cell encapsulation are disclosed in numerous U.S. Patents; among these are Nos. 4,353,888; 4,409,311; 4,673,566; 4,744,933; 4,798,786; 4,803,168; 4,892,538; 5,011,472; 5,158,881; 5,182,111; 5,283,187; 5,474,547; 5,498,401 (which is particularly directed to the encapsulation of bacterial and yeast cells in chitosan); 5,550,050; 5,573,934; 5,578,314; 5,620,883; 5,626,561; 5,653,687; 5,686,115; 5,693,513; and 5,698,413, the contents of which are fully incorporated by reference herein. Typically required for the successful culture of encapsulated cells is a selectively-permeable outer covering or 'skin' which is biocompatible (i.e., tolerated by both the encapsulated cells and the recipient host), and, optionally, a matrix in- or upon which cells are distributed such that the matrix provides structural support and a substrate to which anchorage-dependent cells may attach themselves. As relates to encapsulation devices applicable to use in the invention, the term "selectively-permeable" refers to materials comprising openings through which small molecules (including molecules of up to about 50,000 M.W. - 100,000 M.W.) may pass, but from which larger molecules, such as antibodies (approximately 150,000 M.W.), are excluded. Suitable covering materials include, but are not limited to, porous and/or

polymeric materials such as polyaspartate, polyglutamate, polyacrylates (e.g., acrylic copolymers or RL®, Monsanto Corporation), polyvinylidene fluoride, polyvinylidienes, polyvinyl chloride, polyurethanes, polyurethane isocyanates, polystyrenes, polyamides, cellulose-based polymers (e.g. cellulose acetates and cellulose nitrates), polymethyl-acrylate, polyalginate, polysulfones, polyvinyl alcohols, polyethylene oxide, polyacrylonitriles and derivatives, copolymers and/or mixtures thereof, stretched polytetrafluoroethylene (U.S. Pat. Nos. 3,953,566 and 4,187,390, both incorporated herein by reference), stretched polypropylene, stretched polyethylene, porous polyvinylidene fluoride, woven or non-woven collections of fibers or yarns, such as "Angel Hair" (Anderson, Science, 246: 747-749; Thompson et al., 1989, Proc. Natl. Acad. Sci. U.S.A., 86: 7928-7932), fibrous matrices (see U.S. Pat. No. 5,387,237, incorporated herein by reference), either alone or in combination, or silicon-oxygen-silicon matrices (U.S. Patent No. 5,693,513). Polylysine having a molecular weight of 10,000 to 30,000, preferably 15,000 to 25,000 and most preferably 17,000 is also of use in the invention (see U.S. Patent No. 4,673,566). Alternatively, the matrix material, comprising the transfected cells of the invention, is exposed to conditions that induce it to form its own outer covering, as discussed below.

As described in U.S. Patent No. 5,626,561, the selective permeability of such a covering may be varied by impregnating the void spaces of a porous polymeric material (e.g., stretched polytetrafluoroethylene) with a hydrogel material. Hydrogel material can be impregnated in substantially all of the void spaces of a porous polymeric material or in only a portion of the void spaces. For example, by impregnating a porous polymeric material with a hydrogel material in a continuous band within the material adjacent to and/or along the interior surface of a porous polymeric material, the selective permeability of the material is varied sharply from an outer cross-sectional area of the material to an inner cross-sectional area of the material. The amount and composition of hydrogel material impregnated in a porous polyhmeric material depends in large part on the particular porous polymeric material used to encapsulate cells for transplant. Examples of suitable hydrogel materials

include, but are not limited to, HYPAN® Structural Hydrogel (Hymedix International, Inc.; Dayton, NJ), non-fibrogenic alginate, as taught by Dorian in PCT/US93/05461, which is incorporated herein by reference, agarose, alginic acid, carrageenan, collagen, gelatin, polyvinyl alcohol, poly(2-hydroxyethyl methacrylate), poly(N-vinyl-2-pyrrolidone) or gellan gum, either alone or in combination. The matrix typically has a high surface-area:volume ratio, comprising pores or other spaces in- or on which cells may grow and through which fluids may pass; in addition, suitable matrix materials are stable following transplantation into a recipient organism. Preferably, the matrix comprises an aggregation of multiple particles, fibers or laminae. Alternatively, a matrix may comprise an aqueous solution, such as a physiological buffer or body fluid from the recipient organism (see U.S. Patent No. 5,011,472). Suitable matrix materials include liquid, gelled, polymeric, co-polymeric or particulate formulations of aminated glucopolysachharides (e.g., deacetylated chitin, or "chitosan", which is prepared from the pulverized shells of crabs or other crustaceans, and is commercially available as a dry powder; Cat. # C 3646, Sigma, St. Louis, MO), alginate (U.S. Patent No. 4,409,331), poly-β-1→5-N-acetylglucosamine (p-GlcNAc) polysaccharide species (either alone of formulated as co-polymer with collagen; see U.S. Patent No. 5,686,115), reconstituted extracellular matrix preparations (e.g. Matrigel®; Collaborative Research, Inc, Lexington, MA; Babensee et al., 1992, J. Biomed. Matr. Res., 26: 1401), proteins, polyacrylamide, agarose and others.

Methods by which cells become encapsulated using such materials are both numerous and varied. Encapsulation devices comprising a semi-permeable membrane material, as described above, may be pre-formed, filled with cells (e.g. by injection or other manual means) and then sealed (U.S. Patent Nos. 4,892,538; 5,011,472; 5,626,56; and 5,653,687); such sealing may be effectively permanent (e.g. by the use of heat-sealing), semi-permanent (e.g. by the use of a biocompatible adhesive, such as an epoxy, which will not dissolve or degrade in an aqueous environment) or temporary (e.g. by the use of a removable cap or plug, or by shutting of a valve or stopcock). Methods of permanent and semi-permanent sealing are disclosed in U.S. Patent No.

5,653,687. As an alternative to the use of a pre-formed, semi-permeable cell reservoir, methods by which cells suspended in matrix material and the substance which is to form the outer covering of the encapsulation device are co-extruded under conditions which cause the cell/matrix mixture, which may be in liquid or semi-liquid (i.e., gelled) form to be encased in a continuous tube of the semi-permeable polymer, which either forms, or becomes crosslinked, under the extrusion conditions; such an extrusion procedure may lead to the formation of capsules which have only one cell reservoir (U.S. Patent No. 5,283,187) or which are divided into multiple, discrete compartments (U.S. Patent No. 5,158,881). As an alternative to both types of procedure, a liquid or semi-liquid (i.e., gelled) cell/matrix mixture droplet is suspended either in an agent which induces 'curing' or crosslinking of the outer layer of matrix material to form a semi-permeable barrier (U.S. Patent Nos. 4,798,786 and 5,489,401) or in a solution of polymeric material (or monomers thereof), which will polymerize and/or crosslink upon contact with the cell/matrix droplet such that a semi-permeable membrane is deposited thereon (U.S. Patent Nos. 4,353,888; 4,673,566; 4,744,933; 5,620,883; and 5,693,513).

One of skill in the art is well able to select the appropriate matrix and semi-permeable membrane materials and to construct a cell-encapsulation device as described above.

Implantation of such a device is achieved surgically, *via* standard techniques, to a site at or near the anatomical location to which the product encoded by the gene on the gene of interest is to be delivered, as is deemed safest and most expedient. Such a device may take a convenient shape, including, but not limited to, that of a sphere, pellet or other capsule shape, disk, rod or tube; often, the shape of the device is determined by its method of synthesis. For example, one which is formed by co-extrusion of a cell suspension and a polymeric covering material is typically tubular, while one formed by the deposition of a covering on droplets comprising cells in matrix material might be spherical. As discussed above, the number of cells which must be implanted (and, therefore, encapsulated) is dependent upon the requirements of the recipient organism for the product of the transfected gene. The encapsulation

devices described above are typically small (most usefully, 10μm to 1mm in diameter, so as to permit efficient diffusion of substances back and forth between the outer covering and the cells most deeply embedded in the matrix), and it is contemplated that such devices may carry between 10 and 10^{10} cells each. Should the need for larger numbers of cells be anticipated, a plurality (2, 10 or even 100 or more) of such *in vivo* culturing devices may be made and implanted in a given recipient organism.

An encapsulated cell device may be intended for permanent installation; alternatively, retrieval of the device may be desirable, whether to terminate delivery of the product of the gene of interest to the recipient organism at the discretion of one of skill in the art, such as a physician (who must determine on a case-by-case basis the length of time for which a given cell implant is beneficial to the recipient organism) or to replenish the device with fresh cells after long-term use (i.e. months to years). To the latter end, an implantation device may usefully comprise a retrieval aid, such as a guidewire, and a cap or other port, such as may be opened and re-sealed in order to gain access to the cell reservoir, both as described in U.S. Patent No. 4,892,538.

Live cultures of encapsulated cells have been used successfully to deliver gene products to tissues of a recipient animal. U.S. Patent No. 4,673,566 discloses successful maintenance of normal blood sugar levels in a diabetic rat into which encapsulated rat islet of Langerhans cells were implanted; two administrations of 3,000 cells each together were effective for six months, while a single dose of 1,000 cells was effective for two months.

Similarly, heterospecific transplantation of encapsulated islet cells has been demonstrated to treat diabetes successfully (dog islet cells to a mouse recipient, U.S. Patent No. 5.578,314; porcine islet cells to a mouse recipient, Sun et al., 1992, ASAIO J., 38: 124). It is believed that such an approach is promising for the clinical treatment of diabetes mellitus in humans (Calafiore, 1992, ASAIO J., 38: 34).

It is contemplated that these techniques, which have been applied successfully to untransfected cells, may be utilized advantageously with cells that are transfected with therapeutic nucleic acid molecules of use in the invention.

e. Assay of efficacy of transplanted cells in a recipient organism

The efficacy of the transfected cells so administered and their subsequent maintenance in the recipient host may be assayed either by monitoring the activity of a marker gene, which may additionally be comprised by the transfected construct, or by the direct measurement of either the product (e.g. a protein) encoded by the gene of interest or the reduction in the levels of a protein the production of which it (an antisense message or ribozyme) is designed to inhibit. The assays can be performed using conventional molecular and biochemical techniques, such as are known to one skilled in the art, or may comprise histological sampling (*i.e.*, biopsy) and examination of tranplanted cells or organs.

In addition to direct measurements of protein or nucleic acid levels in blood or target tissues encoded by the gene of interest borne by the vector in transfected/transplanted cells, it is possible to monitor changes in the disease state in patients receiving gene transfer *via* transplantation of cells in which the gene of interest is maintained and compare them to the progression or persistence of disease in patients receiving comparable cells transfected with vector constructs lacking the gene of interest.

Other Dosages and Modes of Administration

A patient that is subject to a disease state which is associated with a frameshift mutation may be treated in accordance with the invention, as described above, via in vivo, ex vivo or in vitro methods. For example in in vivo treatments, a ribozyme or a nucleic acid vector encoding a ribozyme, a wild-type RNA, an antisense RNA or DNA or a sequence encoding the antisense RNA, or a wild-type version of a hybrid wild-type/nonsense protein, can be administered to the patient, preferably in a pharmaceutically acceptable delivery vehicle and a biologically compatible solution, by ingestion, injection, inhalation or any number of other methods. The dosages administered will vary from patient to patient; an "effective dose" will be determined by the level of enhancement of function of the transferred genetic material balanced against any risk of deleterious side effects. Monitoring gene expression and/or the

presence or levels of the encoded mutant RNA or protein or its corresponding "sense" protein will assist in selecting and adjusting the dosages administered. Generally, a composition including a nucleoprotein such as a ribozyme will be administered in single or multiple doses, as determined by the physician, in the range of 50 ug - 1 mg, or within the range of 100 ug - 500 ug. A composition including an oligonucleotide will be administered in a single dose in the range of 5 ng - 10 ug, or within the range of 100 ng - 500 ng. A composition including a wild-type RNA or a vector will be administered in a single dose in the range of 10 ng - 100 ug/kg body weight, preferably in the range of 100 ng - 10 ug/kg body weight, such that at least one copy of the sequence is delivered to each target cell. A composition including a protein, e.g., a wild-type version of a hybrid wild-type/nonsense protein, will be administered in single or multiple doses, as determined by the physician, in the range of 10 ug - 1 mg, or within the range of 100 ug - 50 ug. Any of the above dosages may be administered according to the body weight of the patient, as determined by the physician.

Ex vivo transduction is also contemplated within the present invention. Cell populations can be removed from the patient or otherwise provided, transduced with a vector in accordance with the invention, then reintroduced into the patient. The number of cells reintroduced into the patient will depend upon the efficiency of vector transfer, and will generally be in the range of 10^4 - 10^6 transduced cells/patient.

The cells targeted for ex vivo gene transfer in accordance with the invention include any cells to which the delivery of the vector is desired, for example, neuronal cells or stem cells.

Protein, nucleic acid, or cells administered according to the invention is preferably administered in admixture with a pharmaceutically acceptable carrier substance, e.g., magnesium carbonate, lactose, or a phospholipid to form a micelle, the carrier and protein, nucleic acid or cell together can form a therapeutic composition, e.g., a pill, tablet, capsule or liquid for oral administration to the mammal. Other forms of compositions are also envisioned, e.g., a liquid capable of being administered nasally as drops or spray, or a liquid capable of intravenous, parenteral, subcutaneous, or intraperitoneal administration. The substance administered may be in the form of a

biodegradable sustained release formulation for intramuscular administration. For maximum efficacy, where zero order release is desirable, e.g., an implantable or external pump, e.g., an Infusaid™ pump (Infusaid Corp, MA), may be used.

Kits Useful According to the Invention

The invention encompasses kits for diagnosis or treatment of diseases according to the invention.

A diagnostic kit includes suitable packaging materials and one or more of the following reagents: a nucleic acid probe as defined hereinabove, and optionally means for detecting the probe when bound to its complementary sequences. For example, the nucleic acid probe may be labeled, e.g., radiolabeled, fluorescently labeled, etc., or may be detected via indirect labeling techniques, e.g., using a biotin/avidin system, well known in the art.

A diagnostic system, preferably in kit form, comprises yet another embodiment of this invention. This system is useful for assaying the presence of a hybrid wild-type/nonsense protein or its derivative in cells by the formation of an immune complex. This system includes at least one package that contains an antibody of this invention. Optionally, a kit also may include a positive tissue sample control.

Antibodies are also utilized along with an "indicating group" also sometimes referred to as a "label". The indicating group or label is utilized in conjunction with the antibody as a means for determining whether an immune reaction has taken place, and in some instances for determining the extent of such a reaction.

The terms "indicating group" or "label" are used herein to include single atoms and molecules that are linked to the antibody or used separately, and whether those atoms or molecules are used alone or in conjunction with additional reagents. Such indicating groups or labels are themselves well-known in immunochemistry and constitute a part of this invention only insofar as they are utilized with otherwise novel antibodies, methods and/or systems.

For example, an antigen-specific antibody or antibody fragment is detectably labeled by linking the same to an enzyme and use it in an EIA, or enzyme-linked

immunosorbent assay (ELISA). This enzyme, in turn, when later exposed to a substrate in such a manner as to produce a chemical moiety which can be detected, for example, by spectrophotometric, flourometric or, most preferably, by visual means. The substrate may be a chromogenic substrate which generates a reaction product visible to the naked eye.

Enzymes which can be used to detectably label the binding protein which is specific for the desired detectable mutant protein, include, but are not limited to, alkaline phosphatase, horseradish peroxidase, glucose-6-phosphate dehydrogenase, staphylococcal nuclease, delta-V-steroid isomerase, yeast alcohol dehydrogenase, alpha-glycerophosphate dehydrogenase, triose phosphate isomerase, asparaginase, ribonuclease, urease, catalase, glucose-6-phosphate dehydrogenase, glucoamylase and acetylcholinesterase.

By radioactively labeling the binding protein, for example, the antibody, it is possible to detect the antigen bound to a solid support through the use of a radioimmunoassay (RIA). The radioactive isotope can be detected by such means as the use of a gamma counter or a scintillation counter or by autoradiography. Isotopes which are particularly useful for the purpose of the present invention are: ^{3}H, ^{131}I, ^{14}C, and preferably ^{125}I.

It is also possible to label the first or second binding protein with a fluorescent compound. When the fluorescently labeled antibody is exposed to light of the proper wave length, its presence can then be detected due to fluorescence. Among the most commonly used fluorescent labelling compounds are fluorescein isothiocyanate, rhodamine, phycoerythrin, phycocyanin, allophycocyanin, o-phthaldehyde and fluorescamine.

The first or second binding protein also can be detectably labeled by coupling it to a chemiluminescent compound. The presence of the chemiluminescent-tagged antibody is then determined by detecting the presence of luminescence that arises during the course of a chemical reaction. Examples of particularly useful chemiluminescent labeling compounds are luminol, isoluminol, theromatic acridinium ester, imidazole, acridinium salt and oxalate ester.

Likewise, a bioluminescent compound may be used to label the first or second binding protein. Bioluminescence is a type of chemiluminescence found in biological systems in which a catalytic protein increases the efficiency of the chemiluminescent reaction. The presence of a bioluminescent protein is determined by detecting the presence of luminescence. Important bioluminescent compounds for purposes of labeling are luciferin, luciferase and aequorin.

The invention also includes diagnostic reagents for use in the present invention, such as nucleic acid sequences, probes and antibody molecules, and/or positive tissue controls, as described above, and kits including such reagents for use in diagnosing or treating a disease.

An indicating group or label is preferably supplied along with the antibody and may be packaged therewith or packaged separately. Additional reagents such as hydrogen peroxide and diaminobenzidine, and nickel ammonium sulfate may also be included in the system when an indicating group such as HRP is utilized. Such materials are readily available in commerce, as are many indicating groups, and need not be supplied along with the diagnostic system. In addition, some reagents such as hydrogen peroxide decompose on standing, or are otherwise short-lived like some radioactive elements, and are better supplied by the end-user.

OTHER EMBODIMENTS

It will be understood that the invention is described by way of illustration only. Many other embodiments of the present invention in addition to those herein described will be apparent to those skilled in the art from the description herein given without departing from the scope of the present invention as defined in the appended claims.

Table 1

EARLY ONSET
10-20% of total number of AD cases

familial
60% | 90% unknown (54)
| 10% autosomal dominant (6) — 33% PS1, 5% APP, <1% PS2, 60% unknown

40% **sporadic**

LATE ONSET
80-90% of total number of AD cases

familial
40% | 90% unknown (36)
| 10% autosomal dominant* (4)

60% sporadic

Based upon a population-based cross-section study of dementia, 72% of all demented people suffer from Alzheimer's disease (AD), 16% from vascular dementia, 6% from Parkinson's disease dementia and 6% of other dementias (Ott et al., 1995). Early (<65 years) and late (>65 years) onset (EOAD and LOAD) forms of Alzheimer's disease are distinguished. Familial means that AD was observed in relatives of the first degree. This study is based upon Ott et al., 1995; Van Broeckhoven, 1995; Cruts et al., 1998.
In familial EOAD the majority (54%) is not yet linked to a chromosome, whereas 6% is inherited in a an autosomal dominant way and linked to chromosome 1 (PS2, <1%), 14 (PS1, 33%), 19 (APP, 5%), whereas 60% of the autosomal dominant forms is still not linked. In familial LOAD, the majority (90%) is not yet linked to a chromosome, whereas 10% is inherited in an autosomal dominant way. A subset may be linked to chromosome 12 (Pericak-Vance et al., 1997) and ApoE4 nuclear families.
Risk factors: 65% of all EOAD and 25% of all LOAD cases display ApoE4 polymorphism (one or two E4 alleles). ApoE4 data in early onset AD are based upon a study by Van Broeckhoven and Cruts (n = 102 patients). Other risk factors for late onset AD are butyrylcholinesterase and cytochrome c oxidase.

References
Cruts, M. *et al.* (1998) *Hum. Mol. Genetics* 7, 43-51
Davis, R.E. *et al.* (1997) *Proc. Natl. Acad. Sci. USA* 94, 4526-4531
Lehmann, D.J. *et al.* (1997) *Human Mol. Gen.* 6, 1933-1936
Ott, A. *et al.* (1995) *Br. Med. J.* 310, 970-973
Pericak-Vance, M.A. *et al.* (1997) *JAMA* 278, 1237-1241
Van Broeckhoven, C.L. (1995) *Eur. Neurology* 35, 8-19

Table 2
Clinico-pathological information of controls and AD patients.

NBB/ autopsy no.	Age (years)	Sex (m/f)	Dementia duration (years)	GDS	Postmortem delay (h)	Fixation duration (days)	Brain weight (g)	Cause of death
Non-demented controls								
89003	34	m	-	-	<17	1124	1348	empyema of pleura, fibrous pleuritis and fibrous pericarditis, AIDS
81021	43	m	-	-	23	53	1260	non-Hodgkin lymphoma
94119	51	f	-	-	8	41	1156	sepsis
94125	51	m	-	-	6	47	1518	progressive liposarcoma and ileus
88037	58	m	-	-	41	1088	1797	lung carcinoma, massive hemorrhage
90073	65	f	-	-	41	403	1234	pulmonary embolism
90079	72	m	-	-	4	126	1330	myocardial infarction, cardiogenic shock
91026	80	f	-	-	36	65	1205	cardiogenic shock
91027	82	f	-	-	<55	38	1100	myocardial infarction, ventricular fibrillation
90080	85	m	-	-	5	126	1050	cardiac failure, myocardial infarction, coronary sclerosis, lung emphysema
81007	90	f	-	-	12	48	1110	postoperative infecti ns
90083	90	f	-	-	5	143	1040	metabolic acidosis
Alzheimer cases								
89057	40	m	5-6	7	3	28	1410	AD, cachexia
86055	45	m	11	>4	4	3640	1130	AD, cachexia
90102	49	m	6	7	4	33	1426	AD, epilepsy
91092	54	f	5	6	3	78	1055	AD, cachexia
85013	56	f	4-5	7	22	48	1180	AD, bronchopneumonia, dehydration
92054	61	m	3	6	6	30	1180	AD, renal insufficiency
88073	66	m	±15	7	3	30	1270	AD, ischemic cerebral stroke, cachexia, sepsis
83002 cachexia	70	f	12	7	13	34	780	AD, status epilepticus,
93047	70	m	12	7	4	125	1325	AD, urinary tract infection
91094	73	f	11	7	4	66	1106	AD, dehydration, circulation failure
90118	77	m	7	7	4	75	1168	AD, pneumonia
93044	77	m	>5	na	4	127	1095	AD, bronchial pneumonia
93087	81	m	6	6	4	66	1088	AD, bacterial infection
90015	81	f	6	6	3	28	1020	decompensatio cordis
93045	83	f	14	7	5	127	1005	AD, cachexia, urinary tract infection
91081	85	f	4	4	4	39	1060	AD, metastasis digestive tract
88028	85	f	5	4	2	180	1020	AD, hypovolemic shock
90117	86	m	10	7	4	77	1303	AD, uraemia
91086	88	m	4	5	5	75	1058	AD, decompensatio cordis
86002	90	f	>8	>5	3	38	1060	AD, dehydration
93048	92	f	3	7	4	124	896	AD, cachexia, uraemia
Downs' syndr me								
93162	54	f**	11	na	<17	614	730	DS, bronchopn umonia
92080	58	f**	8	7	10	140	712	DS, epil psia, pn umonia, decompensatio cordis
89005	59	f**	5	7	5	29	812	DS, cardiac arrest
93161	62	f**	11	na	17	585	1100	DS, pneum nia
96015	63	f**	-	-	24	87	980	DS, cardiac-respiratory, insufficiency

94058	64	m***	3	7	7	47	875	DS, dehydration, pneumonia
93028	67	f**	8	7	11	104	859	DS, pneumonia

NBB = Netherlands Brain Bank, **/*** = karyotype 47XX21/47XY21, na = not available, GDS = Global deterioration scale (B. Reisberg, F.H. Ferris, M.J. DeLeon and T. Crook, *Am. J. Psych.* 139, 1136, 1982). [1]non-demented

Table 3 Immunoreactivities in the human frontal cortex (Brodman area 11) for mutant β amyloid precursor protein and ubiquitin-B, the mRNA of which is expressed in the +1 reading frame (βAPP^{+1} and Ubi-B^{+1}). Tissues were obtained fr m controls and neuropathologically confirmed Alzheimer and Down syndrome cases.

NBB autopsy no.	age (years)	sex (m/f)	neuropatho-logical state* plaques	tangles	βAPP^{+1}	Ubi-B^{+1}
Non-demented controls						
89003	34	m	-	-	-	-
81021	43	m	-	-	-	-
94119	51	f	-	-	-	-
94125	51	m	-	-	-	-
88037	58	m	-	-	-	-
90073	65	f	-	-	-	-
90079	72	m	+[a]	-	-	-
91026	80	f	-	-	-	-
91027	82	f	-	-	-	-
90080	85	m	+[b]	+[a]	-	-
81007	90	f	-	+[a]	-	-
90083	90	f	+[b]	-	-	-
% pos. staining					0%	0%
Alzheimer cases						
89057	40	m	+[b]	+[c]	-	+
86055	45	m	+[b]	+[c]	-	+
90102	49	m	+[c]	+[c]	-	+
91092	54	f	+[c]	+[c]	-	+
85013	56	f	+[b]	+[b]	-	-
92054	61	m	+[c]	+[c]	-	+
88073	66	m	+[c]	+[c]	-	+
83002	70	f	+[c]	+[c]	+	+
93047	70	m	+[c]	+[c]	+	-
91094	73	f	+[c]	+[c]	-	+
90118	77	m	+[c]	+[c]	-	+
93044	77	m	+[b]	+[a]	-	+
93087	81	m	+[a]	+[b]	-	+
90015	81	f	+[b]	+[a]	+	+
93045	83	f	+[a]	+[c]	-	+
91081	85	f	+[a]	+[a]	-	-
88028	85	f	+[b]	+[b]	-	+
90117	86	m	+[b]	+[c]	-	+
91086	88	m	+[a]	+[a]	-	+
86002	90	f	+[c]	+[a]	+	+
93048	92	f	+[b]	-	-	-
% pos. staining					19%	80%
Downs' syndrome						
93162	54	f	+[c]	+[c]	+	+
92080	58	f	+[b]	+[c]	+	+
89005	59	f	+[b]	+[c]	+	+
93161	62	f	+[c]	+[c]	+	+
96015	63	f	-	-	-	-
94058	64	m	+[b]	+[c]	+	+
93028	67	f	+[c]	+[c]	+	+
% pos. staining					86%	86%

NBB = Netherlands Brain Bank, *Number of plaques (all types) [1] and tangles as rev aled by Congo Red and Bodian silver staining: a) few, b) moderate, c) many.

Table 4 Immunoreactivities in the human temporal cortex (Brodman area 38) for mutant β amyloid precursor protein and ubiquitin-B, the mRNA of which is expressed in the +1 reading frame (βAPP^{+1} and Ubi-B $^{+1}$). Tissues were obtained from controls and neuropathologically confirmed Alzheimer cases. Down syndrome patients showed Alzheimer pathology.

NBB autopsy no.	age (years)	sex (m/f)	neuropathological state* plaques	tangles	βAPP^{+1}	Ubi-B^{+1}
Non-demented controls						
89003	34	m	-	-	-	-
81021	43	m	-	-	-	-
94119	51	f	-	-	-	-
94125	51	m	-	-	-	-
88037	58	m	-	-	-	-
90073	65	f	-	-	-	-
90079	72	m	-	-	-	-
91026	80	f	-	-	-	-
91027	82	f	-	-	-	-
90080	85	m	+^{b}	+^{b}	-	+
81007[4]	90	f	+^{a}	+^{a}	-	-
90083	90	f	+^{b}	-	-	-
% pos. staining					0%	8%
Alzheimer cases						
89057	40	m	+^{b}	+^{c}	-	+
86055	45	m	+^{b}	+^{c}	+	+
90102	49	m	+^{c}	+^{c}	-	+
91092	54	f	+^{b}	+^{c}	-	+
85013	56	f	+^{b}	+^{c}	-	+
92054	61	m	+^{b}	+^{c}	+	+
88073	66	m	+^{a}	+^{a}	+	+
83002	70	f	+^{b}	+^{c}	+	+
93047	70	m	+^{c}	+^{c}	+	+
91094	73	f	+^{b}	+^{c}	-	+
90118	77	m	+^{c}	+^{c}	-	+
93044	77	m	+^{a}	+^{b}	-	+
93087	81	m	+^{a}	+^{b}	+	+
90015	81	f	+^{c}	+^{c}	+	+
93045	83	f	+^{b}	+^{c}	+	+
88028	85	f	+^{a}	+^{c}	-	+
91081	85	f	+^{b}	+^{c}	-	+
90117	86	m	+^{c}	+^{c}	-	+
91086	88	m	+^{b}	+^{a}	-	+
86002	90	f	+^{c}	+^{c}	+	+
93048	92	f	+^{c}	+^{a}	-	-
% pos. staining					43%	95%
Downs' syndrome						
93162	54	f	+^{c}	+^{c}	+	+
92080	58	f	+^{b}	+^{c}	+	+
89005	59	f	+^{b}	+^{c}	+	+
93161	62	f	+^{c}	+^{c}	+	+
96015	63	f	+^{a}	-	-	-
94058	64	m	+^{b}	+^{b}	+	+
93028	67	f	+^{c}	+^{c}	+	+
% pos. staining					86%	86%

NBB = Netherlands Brain Bank, *Number of plaques (all types) [1] and tangl s as rev aled by Congo Red and Bodian silver staining: a) few, b) moderate, c) many.

Table 5 Immunoreactivities in the human hippocampus for mutant β amyloid precursor protein and ubiquitin-B, the mRNA of which is expressed in the +1 reading frame (βAPP^{+1} and Ubi-B^{+1}). Tissues were obtained from controls and neuropathologically confirmed Alzheimer and Down syndrome cases.

NBB autopsy no.	age (years)	sex (m/f)	neuropathological state: plaques	neuropathological state: tangles	βAPP^{+1}	Ubi-B^{+1}
Non-demented controls						
89003	34	m	-	-	-	-
81021	43	m	-	-	-	-
94119	51	f	-	-	-	-
94125	51	m	-	-	-	-
88037	58	m	-	+[a]	-	-
90073	65	f	-	-	-	-
90079	72	m	+[b]	+[c]	-	+
91026	80	f	+[b]	+[c]	-	+
91027	82	f	-	+[c]	-	+
90080	85	m	+[b]	+[b]	-	+
81007	90	f	+[b]	+[b]	-	+
90083	90	f	-	+[a]	-	+
% pos. staining					0%	50%
Alzheimer cases						
89057	40	m	+[a]	+[c]	-	+
86055	45	m	+[b]	+[c]	-	+
90102	49	m	+[b]	+[c]	-	+
91092	54	f	+[b]	+[c]	+	+
85013	56	f	+[b]	+[a]	-	+
92054	61	m	+[b]	+[c]	+	+
88073	66	f	+[c]	+[c]	-	+
83002	70	f	+[c]	+[c]	+	+
93047	70	m	+[c]	+[c]	+	-
91094	73	f	+[b]	+[c]	-	+
90118	77	m	+[b]	+[c]	+	+
93044	77	m	+[b]	+[c]	+	+
93087	81	m	+[b]	+[c]	+	+
93045	83	f	+[b]	+[c]	-	+
88028	85	f	+[a]	+[b]	-	+
91081	85	f	+[b]	+[c]	-	+
90117	86	m	+[b]	+[c]	+	+
91086	88	m	+[b]	+[c]	+	+
86002	90	f	+[c]	+[c]	-	+
93048	92	f	+[c]	+[c]	+	+
% pos. staining					50%	95%
Downs' syndrome						
93162	54	f	+[b]	+[c]	+	+
92080	58	f	+[c]	+[c]	+	+
89005	59	f	+[b]	+[c]	+	+
93161	62	f	+[c]	+[c]	+	+
96015	63	f	-	-	-	-
94058	64	m	+[b]	+[c]	+	+
93028	67	f	+[c]	+[c]	-	+
% positive staining					71%	86%

NBB = Netherlands Brain Bank, * Number of plaques (all types) [1] and tangles as revealed by Congo Red and Bodian silver staining: a) few, b) moderate, c) many. In the absence of hippocampal tissue, patient #90015 was not studied.

Table 6 Immunoreactivities in the human frontal and temporal cortex and hippocampus for β amyloid precursor protein and ubiquitin-B of which the mRNA is expressed in the +1 reading frame (resulting in βAPP^{+1} and Ubi-B^{+1} protein). Tissues were obtained from controls and neuropathologically confirmed Alzheimer and Down syndrome cases. Immunoreactivity present in tangles, dystrophic neurites and neuritic plaques of patients is expressed as a percentage of the total number of patients studied.

	Frontal cortex (area 11)		Temporal cortex (area 38)		Hippocampus	
	βAPP^{+1}	Ubi-B^{+1}	βAPP^{+1}	Ubi-B^{+1}	βAPP^{+1}	Ubi-B^{+1}
Non dementing controls[1] (n=12)		0	0	0	8*	0
50*						
Alzheimer's disease[2] (n=21)	19	80	43	95	50	95
Down syndrome[3] (n=7)	86	86	86	86	71	86

[1]young (n=6) and aged (n=6) non-demented controls.
[2]early (<65 years, n=10) and late (>65 years, n=11) onset Alzheimer
[3]One Down syndrome patient (#96015; Tables 2-5) did not show any signs of dementia or neuropathology and was immunonegative for βAPP^{+1} and Ubi-B^{+1}.
Controls were matched for sex, age and postmortem delay.
*in old non-demented patients with age related neuropathology (tangles, plaques)

In 11 Parkinson patients no reaction for βAPP^{+1} or Ubi-B^{+1} was found in the nigrostriatal system, except for one patient who also suffered from Alzheimer's disease.

When the three brain areas studies were taken together, βAPP^{+1} immunoreactive structures were present in 71% and Ubi-B^{+1}immunoreactive structures in 100% of the Alzheimer patients.

Table 7

	BASE PAIRS (CODING SEQUENCE OF LONGEST FORM)	GAGAG MOTIFS EXPECTED NUMBER (1:1024)	GAGAG MOTIFS ACTUAL NUMBER
βAPP	2234	2.2	7
Tau	1096	1.1	-
Ubiquitin B	687	0.7	2
Apolipoprotein E4	953	0.9	-
MAP2b	5475	5.3	11
NF-low (68 K)	582	0.6	3
NF-medium (145 K)	2748	2.7	3
NF-H (200 K)	3063	3.1	2
Presenilin I	1392	1.4	3
Presenilin II	1346	1.3	3
Big Tau	2058	2	5
GFAP	1299	1.3	6
P53	1239	1.2	2
BCL2	717	0.7	1
Semaphorin III	2313	2.3	4
HUPF-I	3351	3.3	3
HMG	327	0.3	1
NSP-A	2268	2.3	2

Table 8

					GAGAG MOTIFS					
NO]	CHROMOSOME	EXON NO.	MOL. WEIGHT (kDa) LONGEST FORM	BASE PAIRS CODING SEQ. LONGEST FORM (total genomic seq.-kb)	EXPECTED NUMBER (1:1024)	ACTUAL NUMBER	EXON(S) FROM WHICH +1 PEPTIDE IS DERIVED	PREDICTED MOL. WEIGHT OF +1 PEPTIDE	+ PEPTIDE	[SEQ
βAPP[1]	21q21.3-q22.05	18	135	2234 (170)	2.2	7	9/10	38	RGRTSSKELA	1
TAU[2]	17q21	15	67	1096 (100)	1.1	-	13	36	HGRLAPARHAS	2
UBIQUITIN[3]	17p11.1-p12	2	26	687 (1.8)	0.7	2	1	11	DHHPGSGAQ	4

References:
1. Yoshikai et al., Gene 87, 257, 1990; Selkoe et al., Proc. Natl. Acad. Sci. USA 85, 7341, 1988; 2. Neve et al., J. Mol. Brain Res. 1, 271, 1986; Andreadis et al., Biochemistry 31, 10626, 1992, in exon 4A of big Tau five GAGAG motifs are present; 3. Baker and Board, Nucl. Acids Res. 15, 443, 1987; Webb et al., Am. J. Hum. Genet. 46, 308, 1990.

Table 9

+1 protein sequences (right) predicted by a dinucleotide deletion in an mRNA molecule encoding for different proteins (left)

βAPP^{+1}	RGRTSSKELA	[SEQ ID NO: 1]
Tau^{+1}	HGRLAPARHAS	[SEQ ID NO: 2]
Ubi-B^{+1}	YADLREDPDRQ	[SEQ ID NO: 3]
Ubi-B^{+1}	GGGAQ	[SEQ ID NO: 5]
Ubi-B^{+1}	RQDHHPGSGAQ	[SEQ ID NO: 4]
Ubi-B^{+1}	YADLREDPDRQDHHPGSGAQ	[SEQ ID NO: 1400]
Apo-E	GAPRLPPAQAA	[SEQ ID NO: 6]
MAP2B^{+1}	KTRFQRKGPS	[SEQ ID NO: 7]
Neurofilament-L^{+1}	PGNRSMGHE	[SEQ ID NO: 8]
Neurofilament-M^{+1}	EAEGGSRS	[SEQ ID NO: 9]
Neurofilament-H^{+1}	VGAARDSRAA	[SEQ ID NO: 10]
Presenilin I^{+1}	HDYPPGGSV	[SEQ ID NO: 11]
Presenilin I^{+1}	SIQKFQV	[SEQ ID NO: 12]
Presenilin II^{+1}	VEKPGERGGR	[SEQ ID NO: 13]
Big Tau^{+1}	PLFGRGHKRG	[SEQ ID NO: 14]
GFAP^{+1}	EDRGDAGWRGH	[SEQ ID NO: 15]
P53^{+1}	QERGASPRAAPREH	[SEQ ID NO: 16]
BCL2^{+1}	RQPGDVAPGGQHRPVDD	[SEQ ID NO: 17]
Semaphorin III^{+1}	AGLLAIPEAK	[SEQ ID NO: 18]
Semaphorin III^{+1}	YVDVYNGGKFS	[SEQ ID NO: 19]
HUPF-I^{+1}	AADERRCHLLHMCGRR	[SEQ ID NO: 20]
HUPF-I^{+1}	QQATEAGQHYQPGSPLHDHSHV	[SEQ ID NO: 21]
HMG^{+1}	PQEAAARTNR	[SEQ ID NO: 22]
NSP-A^{+1}	RSWVHPAPPYQMCLG	[SEQ ID NO: 23]
NSP-A^{+1}	GGSRTHPR	[SEQ ID NO: 24]

Table 10

wt	⇨	altered
561 565 GAG AGG Glu Arg		561 565 GAA CGU Glu Arg
1107 1112 GAG AGG Glu Arg		1107 1112 GAG AGG Glu Arg
1128 1133 CGA GAG Arg Glu		1128 1133 CGU GAA Arg Glu
1149 1157 AUG AGA GAA Met Arg Glu		1149 1157 AUG CGC GAA Met Arg Glu
1266 1271 GAG AGA Glu Arg		1266 1271 GAA CGC Glu Arg

CLAIMS

What is claimed is:

1. A method for the diagnosis of a disease caused by or associated with an RNA molecule having a transcript mutation giving rise to a frameshift mutation comprising:
   i. providing a biological sample from a patient suspected of having or developing said disease; and
   ii. detecting in said sample the presence of a mutant RNA molecule having a frameshift mutation or a protein encoded thereby,

wherein detection is indicative of the disease.

2. The method of claim 1, wherein the frameshift mutation comprises a deletion or an insertion of a nucleotide.

3. The method of claim 2, wherein the frameshift mutation is associated with the nucleotide sequence GAGA or CTCT.

4. The method of claim 3, wherein the frameshift mutation comprises a dinucleotide mutation associated with a nucleotide sequence comprising GAGA or CTCT.

5. The method of claim 3, wherein said sequence comprises GAGAX or CTCTX, where X is one of G, A, T, or C.

6. The method of claim 3, wherein said sequence comprises one of GAGAC, CTCTG, GAGAG or CTCTC.

7. The method of claim 1, wherein the disease is cancer or a neurodegenerative disease.

8. The method of claim 7, wherein the disease is Alzheimer's disease or Downs' Syndrome; frontal lobe

dementia (Pick's Disease); progressive supranuclear palsy (PSP) and other diseases with abundant tau-positive filamentous lesions selected from the group that includes Corticobasal degeneration, Dementia pugilistica, Dementia with tangles only, Dementia with tangles and calcification, Frontotemporal dementias with Parkinsonism linked to chromosome 17, Gertsmann-Strässler-Scheinker disease with tangles, Myotonic dystrophy, Niemann-Pick disease type C, Parkinsonism-dementia complex of Guam, Postencephalitic Parkinsonism and Subacute sclerosing panencephalitis; Parkinson's disease; amyotrophic lateral sclerosis; Huntington's Disease; multiple sclerosis; dementi with Lewy bodies; multisystem atrophy; other inclusion body diseases associated with ubiquitin selected from the group that includes Alexander's disease, Alcoholic liver disease, lichen amyloidosis, and the presence of Marinesco bodies and Hyaline inclusions; Diabetes mellitus type II; and other degenerative diseases.

9. The method of claim 1 wherein the RNA having a frameshift mutation would, if containing a wildtype sequence, encode the β amyloid precursor protein, the Tau protein, ubiquitin, apolipoprotein-E_4 (Apo-E_4), microtubule associated protein II (MAP 2), the neurofilament proteins (L, M, H), presenilin I protein, presenilin II protein, Big Tau, GFAP, P53, BCL2, semaphorin III, HUPF-I, HMG and NSP-A.

10. The method of claim 1 wherein the biological sample comprises body fluid or tissue.

11. The method of claim 10 wherein said body fluid comprises cerebral spinal fluid or blood.

12. The method of claim 10, wherein the tissue comprises skin or nose epithelium.

13. The method of claim 1, wherein the mutant RNA molecule is detected by formation of a nucleic acid duplex wherein a first strand of said duplex comprises a nucleic acid probe having a sequence complementary to part of the mutant RNA molecule encompassing the mutation giving rise to the frameshift mutation, and the second strand of said duplex comprises a nucleic acid sequence of the mutant

RNA molecule which is complementary to said probe.

14. The method of claim 1, wherein the mutant RNA molecule is detected using RT-PCR to reverse transcribe the mutant RNA molecule and then to amplify at least a fragment of the reverse transcribed DNA corresponding to the mutant RNA molecule, the mutant RNA molecule encompassing the mutation giving rise to the frameshift, and then probing for the amplified fragment using a nucleic acid probe having a sequence complementary to part of the reverse transcribed DNA encompassing the mutation giving rise to the frameshift mutation, or by sequencing the amplified fragment.

15. The method of claim 1, wherein the protein encoded by the mutant RNA molecule is detected using an antibody molecule having specificity for the mutant protein and not for the wild-type protein.

16. A method for identifying diseases caused by or associated with an RNA molecule having a transcript mutation giving rise to a frameshift mutation comprising:
    i. providing the sequence of an RNA molecule suspected of being involved in the pathogenesis of a disease;
    ii. identifying the sequence of the mutant protein encoded by the RNA sequence 3'-terminal to a frameshift mutation;
    iii. preparing a probe to the mutant protein or a fragment thereof; and
    iv. probing a biological sample from a patient having the disease and a biological sample from a patient not having the disease,

wherein the presence of said mutant protein in a biological sample from a patient having the disease and the absence of said mutant protein in a biological sample from a patient not having the disease indicates that the presence of the mutant protein in a biological sample is a marker for the disease or susceptibility to the disease.

17. A diagnostic kit for diagnosing a disease caused by or associated with an RNA molecule having a transcript mutation giving rise to a frameshift mutation, the kit comprising:

i. a labeled nucleic acid probe having a sequence complementary to part of the mutant RNA molecule which encompasses the mutation which leads to the frameshift mutation; and

ii. packaging materials therefor.

18. A diagnostic kit for diagnosing a disease caused by or assoeiated with an RNA molecule having a transcript mutation giving rise to a frameshift mutation comprising:

i. a pair of primers for use in an RT-PCR reaction, wherein said pair comprises sequences complementary to sequences on either side of the mutation which gives rise to the frameshift mutation, and reagents necessary for performing an RT-PCR reaction; and

ii. packaging materials therefor.

19. A diagnostic kit for diagnosing a disease caused by or associated with at least one RNA molecule having one or more transcript mutations giving rise to a frameshift mutation comprising:

i. an antibody molecule having specificity for the mutant protein encoded by the mutant RNA and not the wild-type protein; and

ii. packaging materials therefor.

20. A recombinant RNA molecule having a frameshift mutation, as described in of any one of claims 1 to 9.

21. The RNA molecule of claim 20 encoding at least part of the protein sequence designated +1 or +2 shown in any one of Figures 2-19.

22. A mutant protein encoded by the RNA of claim 20 or 21.

23. An immunogenic fragment of the mutant protein of claim 22.

24. The mutant protein of claim 22 or the immunogenic fragment of claim 23, comprising the amino acid

sequence:

RGRTSSKELA	[SEQ ID NO: 1];
HGRLAPARHAS	[SEQ ID NO: 2];
YADLREDPDRQ	[SEQ ID NO: 3];
RQDHHPGSGAQ	[SEQ ID NO: 4];
YADLREDPDRQDHHPGSGAQ	[SEQ ID NO: 1400];
GGGAQ	[SEQ ID NO: 5];
GAPRLPPAQAA	[SEQ ID NO: 6];
KTRFQRKGPS	[SEQ ID NO: 7];
PGNRSMGHE	[SEQ ID NO: 8];
EAEGGSRS	[SEQ ID NO: 9];
VGAARDSRAA	[SEQ ID NO: 10];
HDYPPGGSV	[SEQ ID NO: 11];
SIQKFQV	[SEQ ID NO: 12];
VEKPGERGGR	[SEQ ID NO: 13];
PLFGRGHKRG	[SEQ ID NO: 14];
EDRGDAGWRGH	[SEQ ID NO: 15];
QERGASPRAAPREH	[SEQ ID NO: 16];
RQPGDVAPGGQHRPVDD	[SEQ ID NO: 17];
AGLLAIPEAK	[SEQ ID NO: 18];
YVDVYNGGKFS	[SEQ ID NO: 19];
AADERRCHLLHMCGRR	[SEQ ID NO: 20];
QQATEAGQHYQPGSPLHDHSHV	[SEQ ID NO: 21];
PQEAAARTNR	[SEQ ID NO: 22];
RSWVHPAPPYQMCLG	[SEQ ID NO: 23]; or
GGSRTHPR	[SEQ ID NO: 24].

25. A pharmaceutical composition comprising a ribozyme that selectively cleaves a target RNA having a

GAGA or CTCT admixed with a pharmaceutically acceptable carrier.

26. A pharmaceutical composition comprising a ribozyme that selectively cleaves a target RNA having a GAGA or CTCT and a wild-type analog of an RNA having a GAGA sequence giving rise to a frameshift mutation admixed with a pharmaceutically acceptable carrier.

27. A pharmaceutical composition comprising a wild-type analog of an RNA having a GAGA or CTCT sequence giving rise to a frameshift mutation admixed with a pharmaceutically acceptable carrier.

28. The pharmaceutical composition of claim 27 wherein said wild-type analog of an RNA comprises a nucleotide sequence having third base silent mutations.

29. A pharmaceutical composition comprising a single stranded nucleic acid having a sequence that is complementary to an RNA having one or more GAGA or CTCT mutations giving rise to a frameshift mutation admixed with a pharmaceutically acceptable carrier.

30. A pharmaceutical composition comprising the wild-type analog of a mutant protein in admixture with a pharmaceutically acceptable carrier.

31. A vector comprising an expressible gene encoding a ribozyme that selectively cleaves a target RNA having a GAGA or CTCT.

32. A vector comprising an expressible gene encoding a sequence complementary to an RNA having a GAGA or CTCT mutation giving rise to a frameshift mutation.

33. A host cell containing a vector as described in claim 31 or 32.

34. A method of treatment and/or prevention of a disease caused by or associated with an RNA having a

GAGA or CTCT mutation giving rise to a frameshift mutation, comprising administering the composition of any one of claims 25-30, the vector of claim 31 or 32, or the host cell of claim 33 to a patient suffering from or susceptible to the disease.

35. The use of a vector encoding a ribozyme that selectively cleaves a target RNA having a GAGA or CTCT under the control of a promoter in therapy.

36. The use of a vector encoding a ribozyme under the control of a promoter in the manufacture of a composition for the treatment of a disease caused by or associated with at least one an RNA having one or more GAGA or CTCT mutations giving rise to a frameshift mutation.

37. The use of a vector encoding the sequence complementary to an RNA having one or more GAGA or CTCT mutations giving rise to a frameshift mutation under the control of a promoter in therapy.

38. The use of more than one of the composition of any one of claims 25-30, the vector of claim 31 or 32, or the host cell of claim 33 in any combination in therapy.

39. The use of more than one of the composition of any one of claims 25-30, the vector of claim 31 or 32, or the host cell of claim 33 in any combination in the treatment and/or prevention of a disease caused by or associated with at least one an RNA having one or more GAGA or CTCT mutations giving rise to a frameshift mutation.

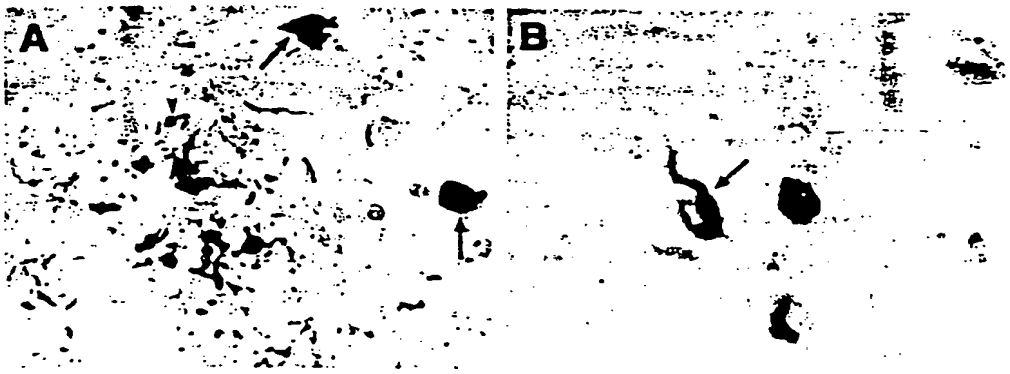

Figure 1

Paraffin section (6 μm thick) of the frontal cortex of a female Alzheimer patient (age 70 years) immunocytochemically incubated with an antibody against a peptide predicted by the +1 reading frame of βAPP. The hallmarks of AD: dystrophic neurites (arrowheads) (A) and tangles (arrows) are clearly visible in cortical layer III. RGRTSSKELA = Amy^{+1} (see Table 9).

Figure 2-1 β amyloid precursor protein

(Linear) MAP of: Seq check: 6510 from: 147 to: 2300

LOCUS HUMAFPA4 3353 bp ss-mRNA PRI 15-JUN-1989
DEFINITION Human amyloid A4 mRNA, complete cds.
ACCESSION Y00264
KEYWORDS amyloid fibril protein; cell surface glycoprotein.
SOURCE human (Homo sapiens).
ORGANISM Homo sapiens . . .

With 1 enzymes: NOTI

September 14, 1993 11:31 ..

```
     ATGCTGCCCGGTTTGGCACTGCTCCTGCTGGCCGCCTGGACGGCTCGGGCGCTGGAGGTA
147  ---+---------+---------+---------+---------+---------+------ 206
     TACGACGGGCCAAACCGTGACGAGGACGACCGGCGGACCTGCCGAGCCCGCGACCTCCAT

a     C  C  P  V  W  H  C  S  C  W  P  P  G  R  L  G  R  W  R  Y  -
b      A  A  R  F  G  T  A  P  A  G  R  L  D  G  S  G  A  G  G  T  -
c    M  L  P  G  L  A  L  L  L  L  A  A  W  T  A  R  A  L  E  V  -

     CCCACTGATGGTAATGCTGGCCTGCTGGCTGAACCCCAGATTGCCATGTTCTGTGGCAGA
207  ---+---------+---------+---------+---------+---------+------ 266
     GGGTGACTACCATTACGACCGGACGACCGACTTGGGGTCTAACGGTACAAGACACCGTCT

a     P  L  M  V  M  L  A  C  W  L  N  P  R  L  P  C  S  V  A  D  -
b      H  *  W  *  C  W  P  A  G  *  T  P  D  C  H  V  L  W  Q  T  -
c    P  T  D  G  N  A  G  L  L  A  E  P  Q  I  A  M  F  C  G  R  -

     CTGAACATGCACATGAATGTCCAGAATGGGAAGTGGGATTCAGATCCATCAGGGACCAAA
267  ---+---------+---------+---------+---------+---------+------ 326
     GACTTGTACGTGTACTTACAGGTCTTACCCTTCACCCTAAGTCTAGGTAGTCCCTGGTTT

a     *  T  C  T  *  M  S  R  M  G  S  G  I  Q  I  H  Q  G  P  K  -
b      E  H  A  H  E  C  P  E  W  E  V  G  F  R  S  I  R  D  Q  N  -
c    L  N  M  H  M  N  V  Q  N  G  K  W  D  S  D  P  S  G  T  K  -

     ACCTGCATTGATACCAAGGAAGGCATCCTGCAGTATTGCCAAGAAGTCTACCCTGAACTG
327  ---+---------+---------+---------+---------+---------+------ 386
     TGGACGTAACTATGGTTCCTTCCGTAGGACGTCATAACGGTTCTTCAGATGGGACTTGAC

a     P  A  L  I  P  R  K  A  S  C  S  I  A  K  K  S  T  L  N  C  -
b      L  H  *  Y  Q  G  R  H  P  A  V  L  P  R  S  L  P  *  T  A  -
c    T  C  I  D  T  K  E  G  I  L  Q  Y  C  Q  E  V  Y  P  E  L  -

     CAGATCACCAATGTGGTAGAAGCCAACCAACCAGTGACCATCCAGAACTGGTGCAAGCGG
387  ---+---------+---------+---------+---------+---------+------ 446
     GTCTAGTGGTTACACCATCTTCGGTTGGTTGGTCACTGGTAGGTCTTGACCACGTTCGCC

a     R  S  P  M  W  *  K  P  T  N  Q  *  P  S  R  T  G  A  S  G  -
b      D  H  Q  C  G  R  S  Q  P  T  S  D  H  P  E  L  V  Q  A  G  -
c    Q  I  T  N  V  V  E  A  N  Q  P  V  T  I  Q  N  W  C  K  R  -
```

Figure 2-2

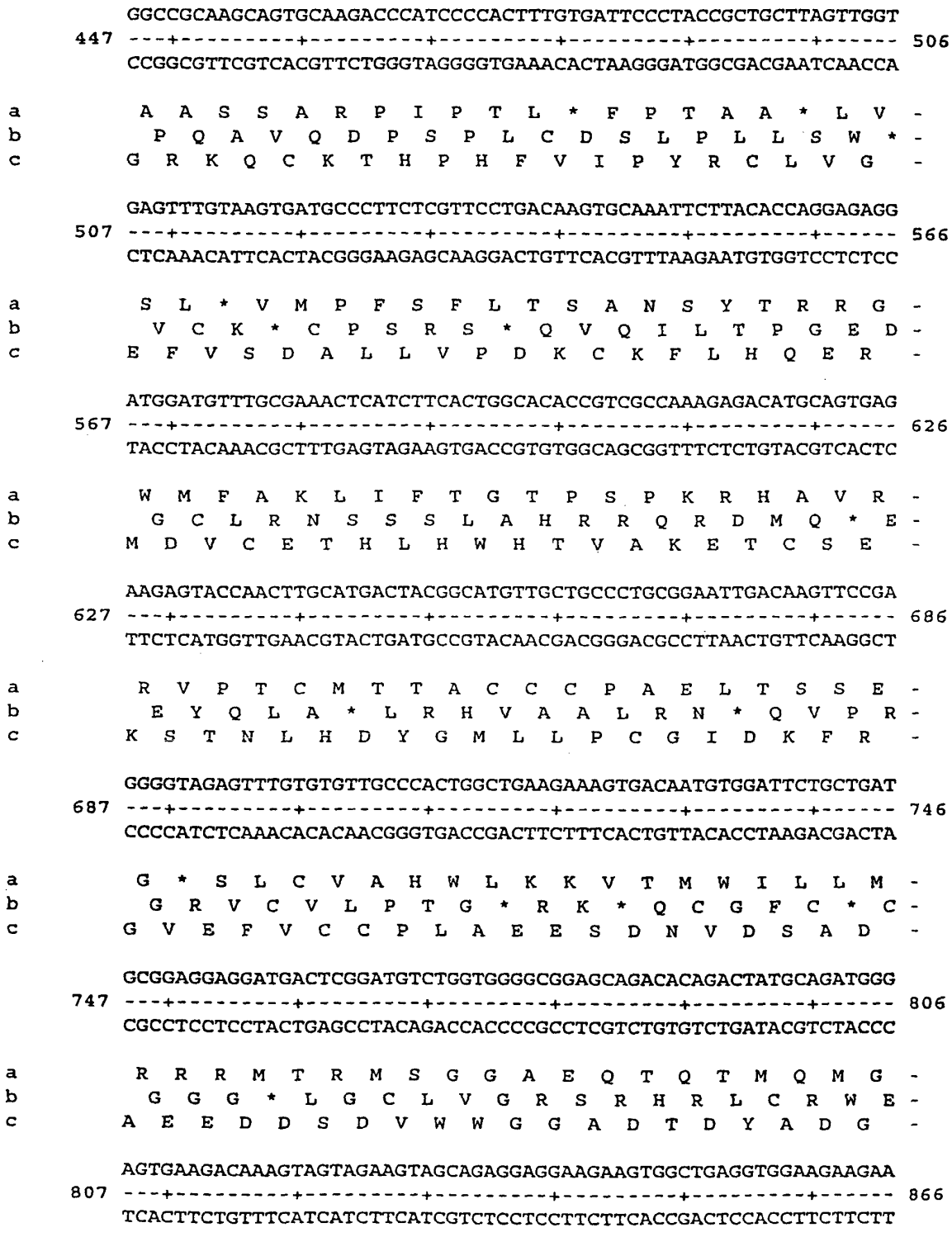
GGCCGCAAGCAGTGCAAGACCCATCCCCACTTTGTGATTCCCTACCGCTGCTTAGTTGGT
447 506
CCGGCGTTCGTCACGTTCTGGGTAGGGGTGAAACACTAAGGGATGGCGACGAATCAACCA
a A A S S A R P I P T L * F P T A A * L V -
b P Q A V Q D P S P L C D S L P L L S W * -
c G R K Q C K T H P H F V I P Y R C L V G -
GAGTTTGTAAGTGATGCCCTTCTCGTTCCTGACAAGTGCAAATTCTTACACCAGGAGAGG
507 566
CTCAAACATTCACTACGGGAAGAGCAAGGACTGTTCACGTTTAAGAATGTGGTCCTCTCC
a S L * V M P F S F L T S A N S Y T R R G -
b V C K * C P S R S * Q V Q I L T P G E D -
c E F V S D A L L V P D K C K F L H Q E R -
ATGGATGTTTGCGAAACTCATCTTCACTGGCACACCGTCGCCAAAGAGACATGCAGTGAG
567 626
TACCTACAAACGCTTTGAGTAGAAGTGACCGTGTGGCAGCGGTTTCTCTGTACGTCACTC
a W M F A K L I F T G T P S P K R H A V R -
b G C L R N S S S L A H R R Q R D M Q * E -
c M D V C E T H L H W H T V A K E T C S E -
AAGAGTACCAACTTGCATGACTACGGCATGTTGCTGCCCTGCGGAATTGACAAGTTCCGA
627 686
TTCTCATGGTTGAACGTACTGATGCCGTACAACGACGGGACGCCTTAACTGTTCAAGGCT
a R V P T C M T T A C C C P A E L T S S E -
b E Y Q L A * L R H V A A L R N * Q V P R -
c K S T N L H D Y G M L L P C G I D K F R -
GGGGTAGAGTTTGTGTGTTGCCCACTGGCTGAAGAAAGTGACAATGTGGATTCTGCTGAT
687 746
CCCCATCTCAAACACACAACGGGTGACCGACTTCTTTCACTGTTACACCTAAGACGACTA
a G * S L C V A H W L K K V T M W I L L M -
b G R V C V L P T G * R K * Q C G F C * C -
c G V E F V C C P L A E E S D N V D S A D -
GCGGAGGAGGATGACTCGGATGTCTGGTGGGGCGGAGCAGACACAGACTATGCAGATGGG
747 806
CGCCTCCTCCTACTGAGCCTACAGACCACCCCGCCTCGTCTGTGTCTGATACGTCTACCC
a R R R M T R M S G G A E Q T Q T M Q M G -
b G G G * L G C L V G R S R H R L C R W E -
c A E E D D S D V W W G G A D T D Y A D G -
AGTGAAGACAAAGTAGTAGAAGTAGCAGAGGAGGAAGAAGTGGCTGAGGTGGAAGAAGAA
807 866
TCACTTCTGTTTCATCATCTTCATCGTCTCCTCCTTCTTCACCGACTCCACCTTCTTCTT

Figure 2-3

```
a          V  K  T  K  *  *  K  *  Q  R  R  K  K  W  L  R  W  K  K  K  -
b            *  R  Q  S  S  R  S  S  R  G  G  R  S  G  *  G  G  R  R  R  -
c          S  E  D  K  V  V  E  V  A  E  E  E  E  V  A  E  V  E  E  E  -

           GAAGCCGATGATGACGAGGACGATGAGGATGGTGATGAGGTAGAGGAAGAGGCTGAGGAA
       867 ---+---------+---------+---------+---------+---------+---------+------ 926
           CTTCGGCTACTACTGCTCCTGCTACTCCTACCACTACTCCATCTCCTTCTCCGACTCCTT

a          K  P  M  M  T  R  T  M  R  M  V  M  R  *  R  K  R  L  R  N  -
b            S  R  *  *  R  G  R  *  G  W  *  *  G  R  G  R  G  *  G  T  -
c          E  A  D  D  D  E  D  D  E  D  G  D  E  V  E  E  E  A  E  E  -

           CCCTACGAAGAAGCCACAGAGAGAACCACCAGCATTGCCACCACCACCACCACCACCACA
       927 ---+---------+---------+---------+---------+---------+---------+------ 986
           GGGATGCTTCTTCGGTGTCTCTCTTGGTGGTCGTAACGGTGGTGGTGGTGGTGGTGGTGT

a          P  T  K  K  P  Q  R  E  P  P  A  L  P  P  P  P  P  P  P  Q  -
b            L  R  R  S  H  R  E  N  H  Q  H  C  H  H  H  H  H  H  H  R  -
c          P  Y  E  E  A  T  E  R  T  T  S  I  A  T  T  T  T  T  T  T  -

           GAGTCTGTGGAAGAGGTGGTTCGAGTTCCTACAACAGCAGCCAGTACCCCTGATGCCGTT
       987 ---+---------+---------+---------+---------+---------+---------+------ 1046
           CTCAGACACCTTCTCCACCAAGCTCAAGGATGTTGTCGTCGGTCATGGGGACTACGGCAA

a          S  L  W  K  R  W  F  E  F  L  Q  Q  Q  P  V  P  L  M  P  L  -
b            V  C  G  R  G  G  S  S  S  Y  N  S  S  Q  Y  P  *  C  R  *  -
c          E  S  V  E  E  V  V  R  V  P  T  T  A  A  S  T  P  D  A  V  -

           GACAAGTATCTCGAGACACCTGGGGATGAGAATGAACATGCCCATTTCCAGAAAGCCAAA
      1047 ---+---------+---------+---------+---------+---------+---------+------ 1106
           CTGTTCATAGAGCTCTGTGGACCCCTACTCTTACTTGTACGGGTAAAGGTCTTTCGGTTT

a          T  S  I  S  R  H  L  G  M  R  M  N  M  P  I  S  R  K  P  K  -
b            Q  V  S  R  D  T  W  G  *  E  *  T  C  P  F  P  E  S  Q  R  -
c          D  K  Y  L  E  T  P  G  D  E  N  E  H  A  H  F  Q  K  A  K  -

           GAGAGGCTTGAGGCCAAGCACCGAGAGAGAATGTCCCAGGTCATGAGAGAATGGGAAGAG
      1107 ---+---------+---------+---------+---------+---------+---------+------ 1166
           CTCTCCGAACTCCGGTTCGTGGCTCTCTCTTACAGGGTCCAGTACTCTCTTACCCTTCTC

a          R  G  L  R  P  S  T  E  R  E  C  P  R  S  *  E  N  G  K  R  -
b            E  A  *  G  Q  A  P  R  E  N  V  P  G  H  E  R  M  G  R  G  -
c          E  R  L  E  A  K  H  R  E  R  M  S  Q  V  M  R  E  W  E  E  -

           GCAGAACGTCAAGCAAAGAACTTGCCTAAAGCTGATAAGAAGGCAGTTATCCAGCATTTC
      1167 ---+---------+---------+---------+---------+---------+---------+------ 1226
           CGTCTTGCAGTTCGTTTCTTGAACGGATTTCGACTATTCTTCCGTCAATAGGTCGTAAAG

a          Q  N  V  K  Q  R  T  C  L  K  L  I  R  R  Q  L  S  S  I  S  -
b            R  T  S  S  K  E  L  A  *  S  *  *  E  G  S  Y  P  A  F  P  -
c          A  E  R  Q  A  K  N  L  P  K  A  D  K  K  A  V  I  Q  H  F  -
```

Figure 2-4

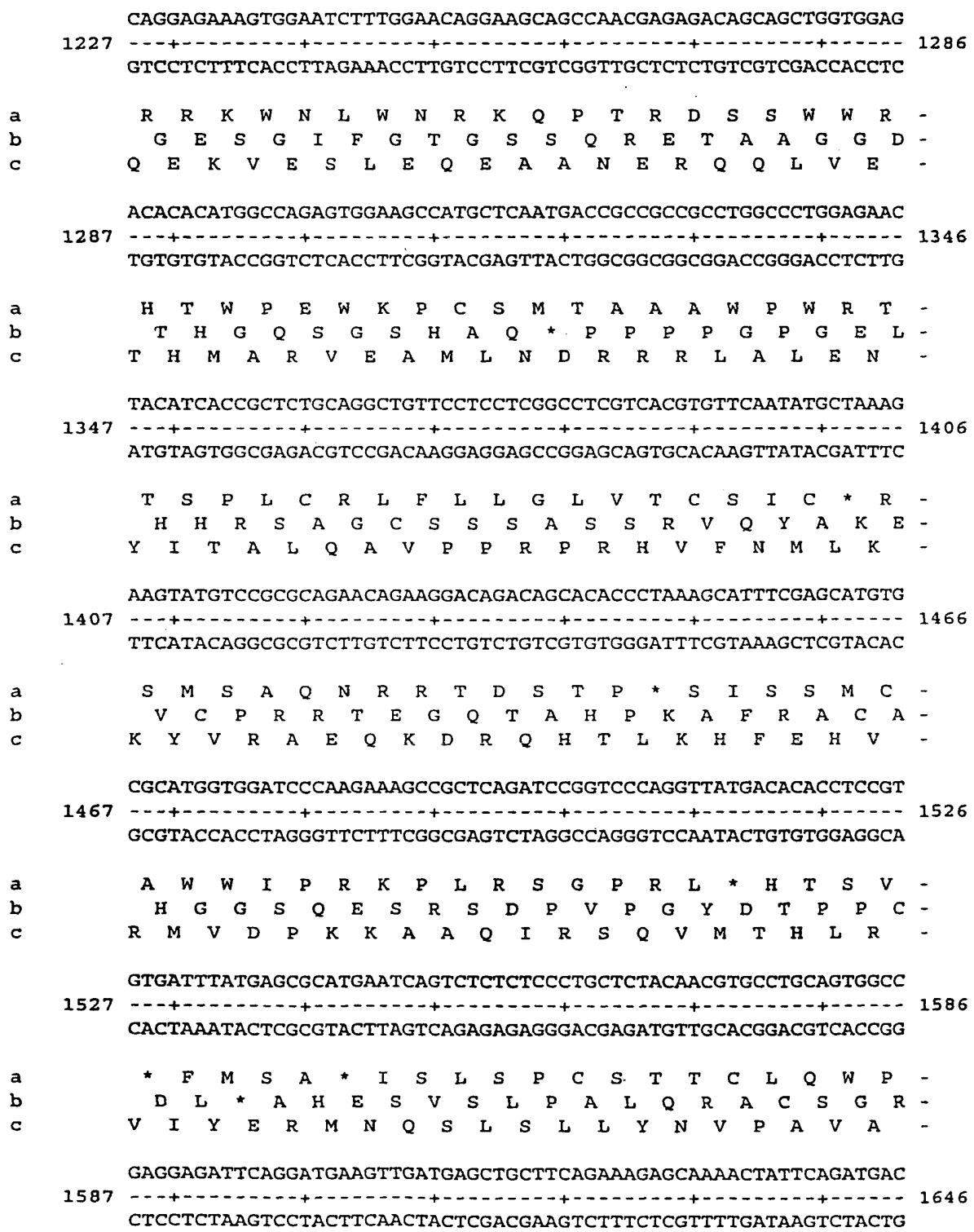
CAGGAGAAAGTGGAATCTTTGGAACAGGAAGCAGCCAACGAGAGACAGCAGCTGGTGGAG
1227 1286
GTCCTCTTTCACCTTAGAAACCTTGTCCTTCGTCGGTTGCTCTCTGTCGTCGACCACCTC
a R R K W N L W N R K Q P T R D S S W W R -
b G E S G I F G T G S S Q R E T A A G G D -
c Q E K V E S L E Q E A A N E R Q Q L V E -
ACACACATGGCCAGAGTGGAAGCCATGCTCAATGACCGCCGCCGCCTGGCCCTGGAGAAC
1287 1346
TGTGTGTACCGGTCTCACCTTCGGTACGAGTTACTGGCGGCGGCGGACCGGGACCTCTTG
a H T W P E W K P C S M T A A A W P W R T -
b T H G Q S G S H A Q * P P P P G P G E L -
c T H M A R V E A M L N D R R R L A L E N -
TACATCACCGCTCTGCAGGCTGTTCCTCCTCGGCCTCGTCACGTGTTCAATATGCTAAAG
1347 1406
ATGTAGTGGCGAGACGTCCGACAAGGAGGAGCCGGAGCAGTGCACAAGTTATACGATTTC
a T S P L C R L F L L G L V T C S I C * R -
b H H R S A G C S S S A S S R V Q Y A K E -
c Y I T A L Q A V P P R P R H V F N M L K -
AAGTATGTCCGCGCAGAACAGAAGGACAGACAGCACACCCTAAAGCATTTCGAGCATGTG
1407 1466
TTCATACAGGCGCGTCTTGTCTTCCTGTCTGTCGTGTGGGATTTCGTAAAGCTCGTACAC
a S M S A Q N R R T D S T P * S I S S M C -
b V C P R R T E G Q T A H P K A F R A C A -
c K Y V R A E Q K D R Q H T L K H F E H V -
CGCATGGTGGATCCCAAGAAAGCCGCTCAGATCCGGTCCCAGGTTATGACACACCTCCGT
1467 1526
GCGTACCACCTAGGGTTCTTTCGGCGAGTCTAGGCCAGGGTCCAATACTGTGTGGAGGCA
a A W W I P R K P L R S G P R L * H T S V -
b H G G S Q E S R S D P V P G Y D T P P C -
c R M V D P K K A A Q I R S Q V M T H L R -
GTGATTTATGAGCGCATGAATCAGTCTCTCTCCCTGCTCTACAACGTGCCTGCAGTGGCC
1527 1586
CACTAAATACTCGCGTACTTAGTCAGAGAGAGGGACGAGATGTTGCACGGACGTCACCGG
a * F M S A * I S L S P C S T T C L Q W P -
b D L * A H E S V S L P A L Q R A C S G R -
c V I Y E R M N Q S L S L L Y N V P A V A -
GAGGAGATTCAGGATGAAGTTGATGAGCTGCTTCAGAAAGAGCAAAACTATTCAGATGAC
1587 1646
CTCCTCTAAGTCCTACTTCAACTACTCGACGAAGTCTTTCTCGTTTTGATAAGTCTACTG

Figure 2-5

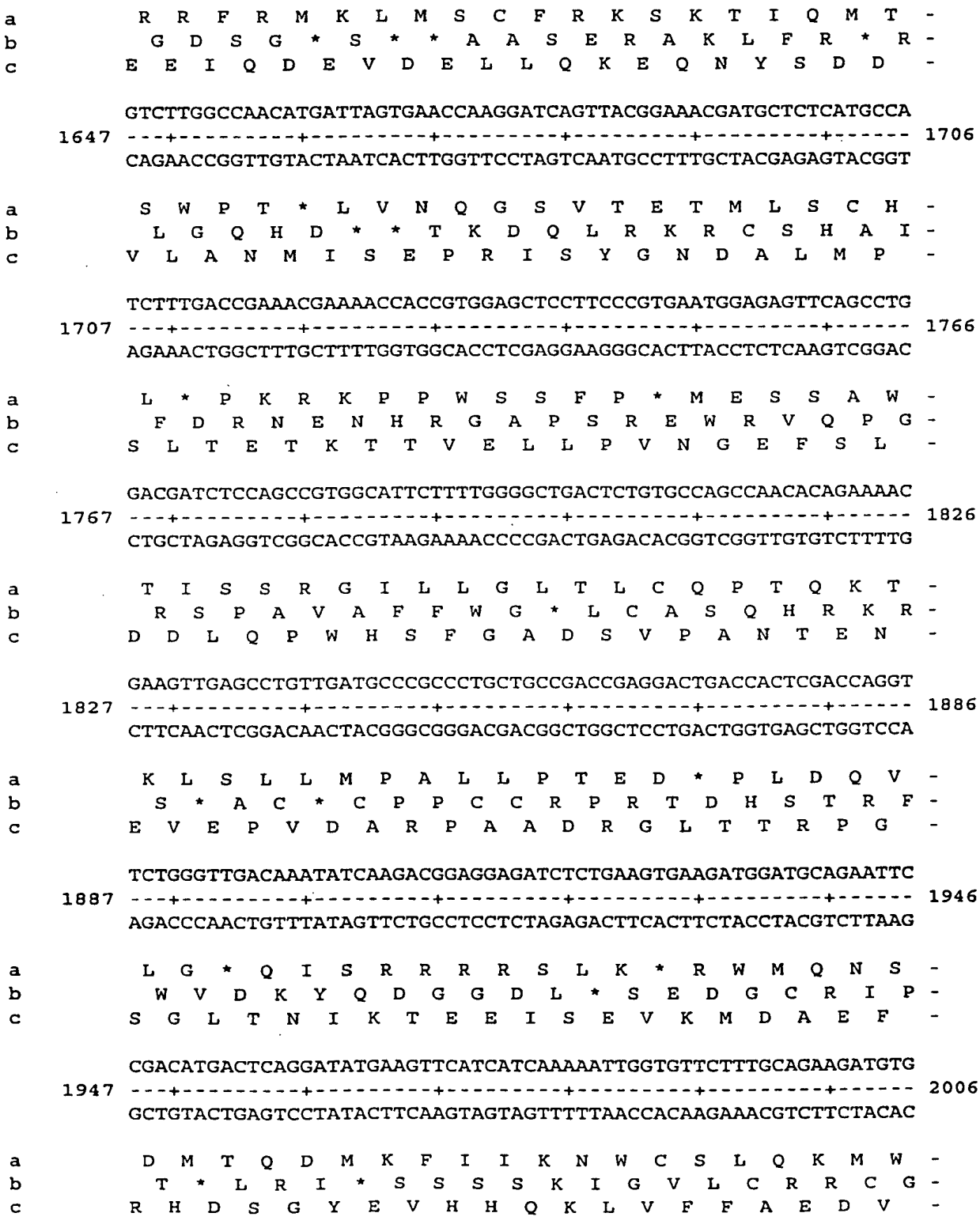
a R R F R M K L M S C F R K S K T I Q M T -
b G D S G * S * * A A S E R A K L F R * R -
c E E I Q D E V D E L L Q K E Q N Y S D D -
GTCTTGGCCAACATGATTAGTGAACCAAGGATCAGTTACGGAAACGATGCTCTCATGCCA
1647 ---+---------+---------+---------+---------+---------+---------+------ 1706
CAGAACCGGTTGTACTAATCACTTGGTTCCTAGTCAATGCCTTTGCTACGAGAGTACGGT
a S W P T * L V N Q G S V T E T M L S C H -
b L G Q H D * * T K D Q L R K R C S H A I -
c V L A N M I S E P R I S Y G N D A L M P -
TCTTTGACCGAAACGAAAACCACCGTGGAGCTCCTTCCCGTGAATGGAGAGTTCAGCCTG
1707 ---+---------+---------+---------+---------+---------+---------+------ 1766
AGAAACTGGCTTTGCTTTTGGTGGCACCTCGAGGAAGGGCACTTACCTCTCAAGTCGGAC
a L * P K R K P P W S S F P * M E S S A W -
b F D R N E N H R G A P S R E W R V Q P G -
c S L T E T K T T V E L L P V N G E F S L -
GACGATCTCCAGCCGTGGCATTCTTTTGGGGCTGACTCTGTGCCAGCCAACACAGAAAAC
1767 ---+---------+---------+---------+---------+---------+---------+------ 1826
CTGCTAGAGGTCGGCACCGTAAGAAAACCCCGACTGAGACACGGTCGGTTGTGTCTTTTG
a T I S S R G I L L G L T L C Q P T Q K T -
b R S P A V A F F W G * L C A S Q H R K R -
c D D L Q P W H S F G A D S V P A N T E N -
GAAGTTGAGCCTGTTGATGCCCGCCCTGCTGCCGACCGAGGACTGACCACTCGACCAGGT
1827 ---+---------+---------+---------+---------+---------+---------+------ 1886
CTTCAACTCGGACAACTACGGGCGGGACGACGGCTGGCTCCTGACTGGTGAGCTGGTCCA
a K L S L L M P A L L P T E D * P L D Q V -
b S * A C * C P P C C R P R T D H S T R F -
c E V E P V D A R P A A D R G L T T R P G -
TCTGGGTTGACAAATATCAAGACGGAGGAGATCTCTGAAGTGAAGATGGATGCAGAATTC
1887 ---+---------+---------+---------+---------+---------+---------+------ 1946
AGACCCAACTGTTTATAGTTCTGCCTCCTCTAGAGACTTCACTTCTACCTACGTCTTAAG
a L G * Q I S R R R R S L K * R W M Q N S -
b W V D K Y Q D G G D L * S E D G C R I P -
c S G L T N I K T E E I S E V K M D A E F -
CGACATGACTCAGGATATGAAGTTCATCATCAAAAATTGGTGTTCTTTGCAGAAGATGTG
1947 ---+---------+---------+---------+---------+---------+---------+------ 2006
GCTGTACTGAGTCCTATACTTCAAGTAGTAGTTTTTAACCACAAGAAACGTCTTCTACAC
a D M T Q D M K F I I K N W C S L Q K M W -
b T * L R I * S S S S K I G V L C R R C G -
c R H D S G Y E V H H Q K L V F F A E D V -

Figure 2-6

```
          GGTTCAAACAAAGGTGCAATCATTGGACTCATGGTGGGCGGTGTTGTCATAGCGACAGTG
     2007 ---+---------+---------+---------+---------+---------+---------+------ 2066
          CCAAGTTTGTTTCCACGTTAGTAACCTGAGTACCACCCGCCACAACAGTATCGCTGTCAC

a          V  Q  T  K  V  Q  S  L  D  S  W  W  A  V  L  S  *  R  Q  *  -
b            F  K  Q  R  C  N  H  W  T  H  G  G  R  C  C  H  S  D  S  D -
c          G  S  N  K  G  A  I  I  G  L  M  V  G  G  V  V  I  A  T  V  -

          ATCGTCATCACCTTGGTGATGCTGAAGAAGAAACAGTACACATCCATTCATCATGGTGTG
     2067 ---+---------+---------+---------+---------+---------+---------+------ 2126
          TAGCAGTAGTGGAACCACTACGACTTCTTCTTTGTCATGTGTAGGTAAGTAGTACCACAC

a          S  S  S  P  W  *  C  *  R  R  N  S  T  H  P  F  I  M  V  W  -
b            R  H  H  L  G  D  A  E  E  E  T  V  H  I  H  S  S  W  C  G -
c          I  V  I  T  L  V  M  L  K  K  K  Q  Y  T  S  I  H  H  G  V  -

          GTGGAGGTTGACGCCGCTGTCACCCCAGAGGAGCGCCACCTGTCCAAGATGCAGCAGAAC
     2127 ---+---------+---------+---------+---------+---------+---------+------ 2186
          CACCTCCAACTGCGGCGACAGTGGGGTCTCCTCGCGGTGGACAGGTTCTACGTCGTCTTG

a          W  R  L  T  P  L  S  P  Q  R  S  A  T  C  P  R  C  S  R  T  -
b            G  G  *  R  R  C  H  P  R  G  A  P  P  V  Q  D  A  A  E  R -
c          V  E  V  D  A  A  V  T  P  E  E  R  H  L  S  K  M  Q  Q  N  -

          GGCTACGAAAATCCAACCTACAAGTTCTTTGAGCAGATGCAGAACTAGACCCCCGCCACA
     2187 ---+---------+---------+---------+---------+---------+---------+------ 2246
          CCGATGCTTTTAGGTTGGATGTTCAAGAAACTCGTCTACGTCTTGATCTGGGGGCGGTGT

a          A  T  K  I  Q  P  T  S  S  L  S  R  C  R  T  R  P  P  P  Q  -
b            L  R  K  S  N  L  Q  V  L  *  A  D  A  E  L  D  P  R  H  S -
c          G  Y  E  N  P  T  Y  K  F  F  E  Q  M  Q  N  *  T  P  A  T  -

          GCAGCCTCTGAAGTTGGACAGCAAAACCATTGCTTCACTACCCATCGGTGTCCA
     2247 ---+---------+---------+---------+---------+---------+---------+ 2300
          CGTCGGAGACTTCAACCTGTCGTTTTGGTAACGAAGTGATGGGTAGCCACAGGT

a          Q  P  L  K  L  D  S  K  T  I  A  S  L  P  I  G  V     -
b            S  L  *  S  W  T  A  K  P  L  L  H  Y  P  S  V  S     -
c          A  A  S  E  V  G  Q  Q  N  H  C  F  T  T  H  R  C  P   -
```

Enzymes that do cut:

NONE

Enzymes that do not cut:

NotI

Figure 3-1 Tau

(Linear) MAP of: Seq check: 9711 from: 38 to: 1096

```
RL;HSTAUA      - Human microtubule-associated protein tau mRNA, complete cds
ID   HSTAUA      standard; RNA; PRI; 1108 BP.
XX
AC   J03778;
XX
DT   04-OCT-1988 (Rel. 17, Created) . . .

 With 1 enzymes: NOTI
                                 September 14, 1993  12:12  ..

          ATGGCTGAGCCCCGCCAGGAGTTCGAAGTGATGGAAGATCACGCTGGGACGTACGGGTTG
       38 --+---------+---------+---------+---------+---------+---------+------- 97
          TACCGACTCGGGGCGGTCCTCAAGCTTCACTACCTTCTAGTGCGACCCTGCATGCCCAAC

a           G  *  A  P  P  G  V  R  S  D  G  R  S  R  W  D  V  R  V  G -
b          M  A  E  P  R  Q  E  F  E  V  M  E  D  H  A  G  T  Y  G  L  -
c           W  L  S  P  A  R  S  S  K  *  W  K  I  T  L  G  R  T  G  W -

          GGGGACAGGAAAGATCAGGGGGGCTACACCATGCACCAAGACCAAGAGGGTGACACGGAC
       98 --+---------+---------+---------+---------+---------+---------+------- 157
          CCCCTGTCCTTTCTAGTCCCCCCGATGTGGTACGTGGTTCTGGTTCTCCCACTGTGCCTG

a           G  Q  E  R  S  G  G  L  H  H  A  P  R  P  R  G  *  H  G  R -
b          G  D  R  K  D  Q  G  G  Y  T  M  H  Q  D  Q  E  G  D  T  D  -
c           G  T  G  K  I  R  G  A  T  P  C  T  K  T  K  R  V  T  R  T -

          GCTGGCCTGAAAGCTGAAGAAGCAGGCATTGGAGACACCCCCAGCCTGGAAGACGAAGCT
      158 --+---------+---------+---------+---------+---------+---------+------- 217
          CGACCGGACTTTCGACTTCTTCGTCCGTAACCTCTGTGGGGGTCGGACCTTCTGCTTCGA

a           W  P  E  S  *  R  S  R  H  W  R  H  P  Q  P  G  R  R  S  C -
b          A  G  L  K  A  E  E  A  G  I  G  D  T  P  S  L  E  D  E  A  -
c           L  A  *  K  L  K  K  Q  A  L  E  T  P  P  A  W  K  T  K  L -

          GCTGGTCACGTGACCCAAGCTCGCATGGTCAGTAAAAGCAAAGACGGGACTGGAAGCGAT
      218 --+---------+---------+---------+---------+---------+---------+------- 277
          CGACCAGTGCACTGGGTTCGAGCGTACCAGTCATTTTCGTTTCTGCCCTGACCTTCGCTA

a           W  S  R  D  P  S  S  H  G  Q  *  K  Q  R  R  D  W  K  R  * -
b          A  G  H  V  T  Q  A  R  M  V  S  K  S  K  D  G  T  G  S  D  -
c           L  V  T  *  P  K  L  A  W  S  V  K  A  K  T  G  L  E  A  M -
```

Figure 3-2

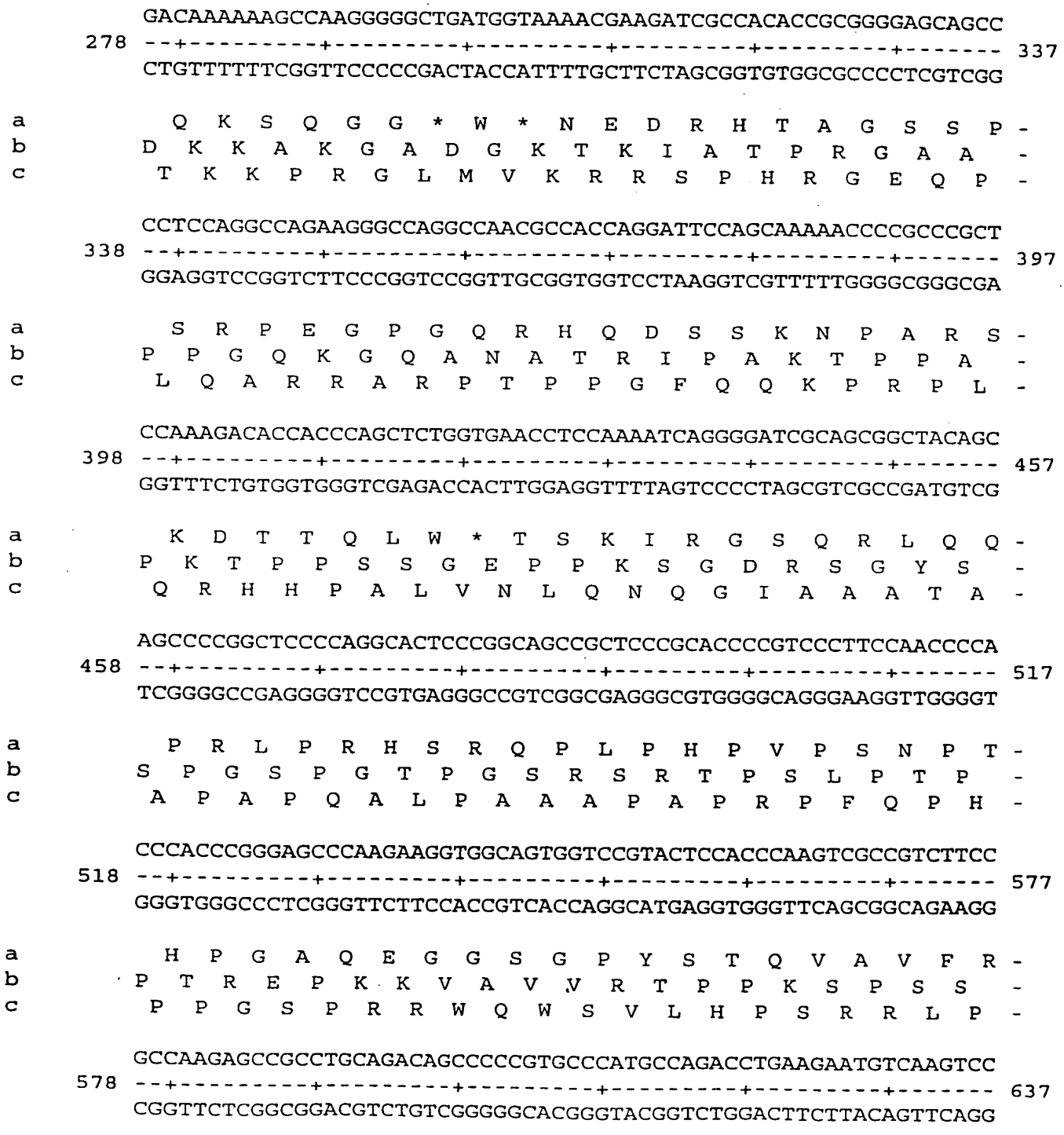
GACAAAAAAGCCAAGGGGGCTGATGGTAAAACGAAGATCGCCACACCGCGGGGAGCAGCC
278
337
CTGTTTTTTCGGTTCCCCCGACTACCATTTTGCTTCTAGCGGTGTGGCGCCCCTCGTCGG
a Q K S Q G G * W * N E D R H T A G S S P -
b D K K A K G A D G K T K I A T P R G A A -
c T K K P R G L M V K R R S P H R G E Q P -
CCTCCAGGCCAGAAGGGCCAGGCCAACGCCACCAGGATTCCAGCAAAAACCCCGCCCGCT
338
397
GGAGGTCCGGTCTTCCCGGTCCGGTTGCGGTGGTCCTAAGGTCGTTTTTGGGGCGGGCGA
a S R P E G P G Q R H Q D S S K N P A R S -
b P P G Q K G Q A N A T R I P A K T P P A -
c L Q A R R A R P T P P G F Q Q K P R P L -
CCAAAGACACCACCCAGCTCTGGTGAACCTCCAAAATCAGGGGATCGCAGCGGCTACAGC
398
457
GGTTTCTGTGGTGGGTCGAGACCACTTGGAGGTTTTAGTCCCCTAGCGTCGCCGATGTCG
a K D T T Q L W * T S K I R G S Q R L Q Q -
b P K T P P S S G E P P K S G D R S G Y S -
c Q R H H P A L V N L Q N Q G I A A A T A -
AGCCCCGGCTCCCCAGGCACTCCCGGCAGCCGCTCCCGCACCCCGTCCCTTCCAACCCCA
458
517
TCGGGGCCGAGGGGTCCGTGAGGGCCGTCGGCGAGGGCGTGGGGCAGGGAAGGTTGGGGT
a P R L P R H S R Q P L P H P V P S N P T -
b S P G S P G T P G S R S R T P S L P T P -
c A P A P Q A L P A A A P A P R P F Q P H -
CCCACCCGGGAGCCCAAGAAGGTGGCAGTGGTCCGTACTCCACCCAAGTCGCCGTCTTCC
518
577
GGGTGGGCCCTCGGGTTCTTCCACCGTCACCAGGCATGAGGTGGGTTCAGCGGCAGAAGG
a H P G A Q E G G S G P Y S T Q V A V F R -
b P T R E P K K V A V V R T P P K S P S S -
c P P G S P R R W Q W S V L H P S R R L P -
GCCAAGAGCCGCCTGCAGACAGCCCCCGTGCCCATGCCAGACCTGAAGAATGTCAAGTCC
578
637
CGGTTCTCGGCGGACGTCTGTCGGGGGCACGGGTACGGTCTGGACTTCTTACAGTTCAGG

Figure 3-3

```
a          Q  E  P  P  A  D  S  P  R  A  H  A  R  P  E  E  C  Q  V  Q  -
b       A  K  S  R  L  Q  T  A  P  V  P  M  P  D  L  K  N  V  K  S  -
c        P  R  A  A  C  R  Q  P  P  C  P  C  Q  T  *  R  M  S  S  P  -

        AAGATCGGCTCCACTGAGAACCTGAAGCACCAGCCGGGAGGCGGGAAGGTGCAAATAGTC
    638 --+---------+---------+---------+---------+---------+---------+------- 697
        TTCTAGCCGAGGTGACTCTTGGACTTCGTGGTCGGCCCTCCGCCCTTCCACGTTTATCAG

a          D  R  L  H  *  E  P  E  A  P  A  G  R  R  E  G  A  N  S  L  -
b       K  I  G  S  T  E  N  L  K  H  Q  P  G  G  G  K  V  Q  I  V  -
c        R  S  A  P  L  R  T  *  S  T  S  R  E  A  G  R  C  K  *  S  -

        TACAAACCAGTTGACCTGAGCAAGGTGACCTCCAAGTGTGGCTCATTAGGCAACATCCAT
    698 --+---------+---------+---------+---------+---------+---------+------- 757
        ATGTTTGGTCAACTGGACTCGTTCCACTGGAGGTTCACACCGAGTAATCCGTTGTAGGTA

a          Q  T  S  *  P  E  Q  G  D  L  Q  V  W  L  I  R  Q  H  P  S  -
b       Y  K  P  V  D  L  S  K  V  T  S  K  C  G  S  L  G  N  I  H  -
c        T  N  Q  L  T  *  A  R  *  P  P  S  V  A  H  *  A  T  S  I  -

        CATAAACCAGGAGGTGGCCAGGTGGAAGTAAAATCTGAGAAGCTTGACTTCAAGGACAGA
    758 --+---------+---------+---------+---------+---------+---------+------- 817
        GTATTTGGTCCTCCACCGGTCCACCTTCATTTTAGACTCTTCGAACTGAAGTTCCTGTCT

a          *  T  R  R  W  P  G  G  S  K  I  *  E  A  *  L  Q  G  Q  S  -
b       H  K  P  G  G  G  Q  V  E  V  K  S  E  K  L  D  F  K  D  R  -
c        I  N  Q  E  V  A  R  W  K  *  N  L  R  S  L  T  S  R  T  E  -

        GTCCAGTCGAAGATTGGGTCCCTGGACAATATCACCCACGTCCCTGGCGGAGGAAATAAA
    818 --+---------+---------+---------+---------+---------+---------+------- 877
        CAGGTCAGCTTCTAACCCAGGGACCTGTTATAGTGGGTGCAGGGACCGCCTCCTTTATTT

a          P  V  E  D  W  V  P  G  Q  Y  H  P  R  P  W  R  R  K  *  K  -
b       V  Q  S  K  I  G  S  L  D  N  I  T  H  V  P  G  G  G  N  K  -
c        S  S  R  R  L  G  P  W  T  I  S  P  T  S  L  A  E  E  I  K  -

        AAGATTGAAACCCACAAGCTGACCTTCCGCGAGAACGCCAAAGCCAAGACAGACCACGGG
    878 --+---------+---------+---------+---------+---------+---------+------- 937
        TTCTAACTTTGGGTGTTCGACTGGAAGGCGCTCTTGCGGTTTCGGTTCTGTCTGGTGCCC

a          D  *  N  P  Q  A  D  L  P  R  E  R  Q  S  Q  D  R  P  R  G  -
b       K  I  E  T  H  K  L  T  F  R  E  N  A  K  A  K  T  D  H  G  -
c        R  L  K  P  T  S  *  P  S  A  R  T  P  K  P  R  Q  T  T  G  -
```

Figure 3-4

```
        GCGGAGATCGTGTACAAGTCGCCAGTGGTGTCTGGGGACACGTCTCCACGGCATCTCAGC
    938 --+---------+---------+---------+---------+---------+-------- 997
        CGCCTCTAGCACATGTTCAGCGGTCACCACAGACCCCTGTGCAGAGGTGCCGTAGAGTCG

a         G  D  R  V  Q  V  A  S  G  V  W  G  H  V  S  T  A  S  Q  Q -
b       A  E  I  V  Y  K  S  P  V  V  S  G  D  T  S  P  R  H  L  S  -
c        R  R  S  C  T  S  R  Q  W  C  L  G  T  R  L  H  G  I  S  A -

        AATGTCTCCTCCACCGGCAGCATCGACATGGTAGACTCGCCCCAGCTCGCCACGCTAGCT
    998 --+---------+---------+---------+---------+---------+-------- 1057
        TTACAGAGGAGGTGGCCGTCGTAGCTGTACCATCTGAGCGGGGTCGAGCGGTGCGATCGA

a         C  L  L  H  R  Q  H  R  H  G  R  L  A  P  A  R  H  A  S  * -
b       N  V  S  S  T  G  S  I  D  M  V  D  S  P  Q  L  A  T  L  A  -
c        M  S  P  P  P  A  A  S  T  W  *  T  R  P  S  S  P  R  *  L -

        GACGAGGTGTCTGCCTCCCTGGCCAAGCAGGGTTTGTGA
   1058 --+---------+---------+---------+------ 1096
        CTGCTCCACAGACGGAGGGACCGGTTCGTCCCAAACACT

a         R  G  V  C  L  P  G  Q  A  G  F  V    -
b       D  E  V  S  A  S  L  A  K  Q  G  L  *    -
c        T  R  C  L  P  P  W  P  S  R  V  C      -
```

Enzymes that do cut:

NONE

Enzymes that do not cut:

NotI

Figure 4-1 Ubiquitin B

(Linear) MAP of: Seq check: 2987 from: 1094 to: 1800

LOCUS HUMYUBG1 2118 bp ds-DNA PRI 15-MAR-1988
DEFINITION Human ubiquitin gene (3 repeats).
ACCESSION X04803
KEYWORDS ubiquitin.
SOURCE human (Homo sapiens).
ORGANISM Homo sapiens . . .

With 1 enzymes: NOTI
September 14, 1993 11:58 ..

```
        ATGCAGATCTTCGTGAAAACCCTTACCGGCAAGACCATCACCCTTGAGGTGGAGCCCAGT
   1094 ------+---------+---------+---------+---------+---------+--- 1153
        TACGTCTAGAAGCACTTTTGGGAATGGCCGTTCTGGTAGTGGGAACTCCACCTCGGGTCA

a         A  D  L  R  E  N  P  Y  R  Q  D  H  H  P  *  G  G  A  Q  * -
b       M  Q  I  F  V  K  T  L  T  G  K  T  I  T  L  E  V  E  P  S  -
c        C  R  S  S  *  K  P  L  P  A  R  P  S  P  L  R  W  S  P  V -

        GACACCATCGAAAATGTGAAGGCCAAGATCCAGGATAAGGAAGGCATTCCCCCCGACCAG
   1154 ------+---------+---------+---------+---------+---------+--- 1213
        CTGTGGTAGCTTTTACACTTCCGGTTCTAGGTCCTATTCCTTCCGTAAGGGGGGCTGGTC

a         H  H  R  K  C  E  G  Q  D  P  G  *  G  R  H  S  P  R  P  A -
b       D  T  I  E  N  V  K  A  K  I  Q  D  K  E  G  I  P  P  D  Q  -
c        T  P  S  K  M  *  R  P  R  S  R  I  R  K  A  F  P  P  T  S -

        CAGAGGCTCATCTTTGCAGGCAAGCAGCTGGAAGATGGCCGTACTCTTTCTGACTACAAC
   1214 ------+---------+---------+---------+---------+---------+--- 1273
        GTCTCCGAGTAGAAACGTCCGTTCGTCGACCTTCTACCGGCATGAGAAAGACTGATGTTG

a         E  A  H  L  C  R  Q  A  A  G  R  W  P  Y  S  F  *  L  Q  H -
b       Q  R  L  I  F  A  G  K  Q  L  E  D  G  R  T  L  S  D  Y  N  -
c        R  G  S  S  L  Q  A  S  S  W  K  M  A  V  L  F  L  T  T  T -

        ATCCAGAAGGAGTCGACCCTGCACCTGGTCCTGCGTCTGAGAGGTGGTATGCAGATCTTC
   1274 ------+---------+---------+---------+---------+---------+--- 1333
        TAGGTCTTCCTCAGCTGGGACGTGGACCAGGACGCAGACTCTCCACCATACGTCTAGAAG

a         P  E  G  V  D  P  A  P  G  P  A  S  E  R  W  Y  A  D  L  R -
b       I  Q  K  E  S  T  L  H  L  V  L  R  L  R  G  G  M  Q  I  F  -
c        S  R  R  S  R  P  C  T  W  S  C  V  *  E  V  V  C  R  S  S -
```

Figure 4-2

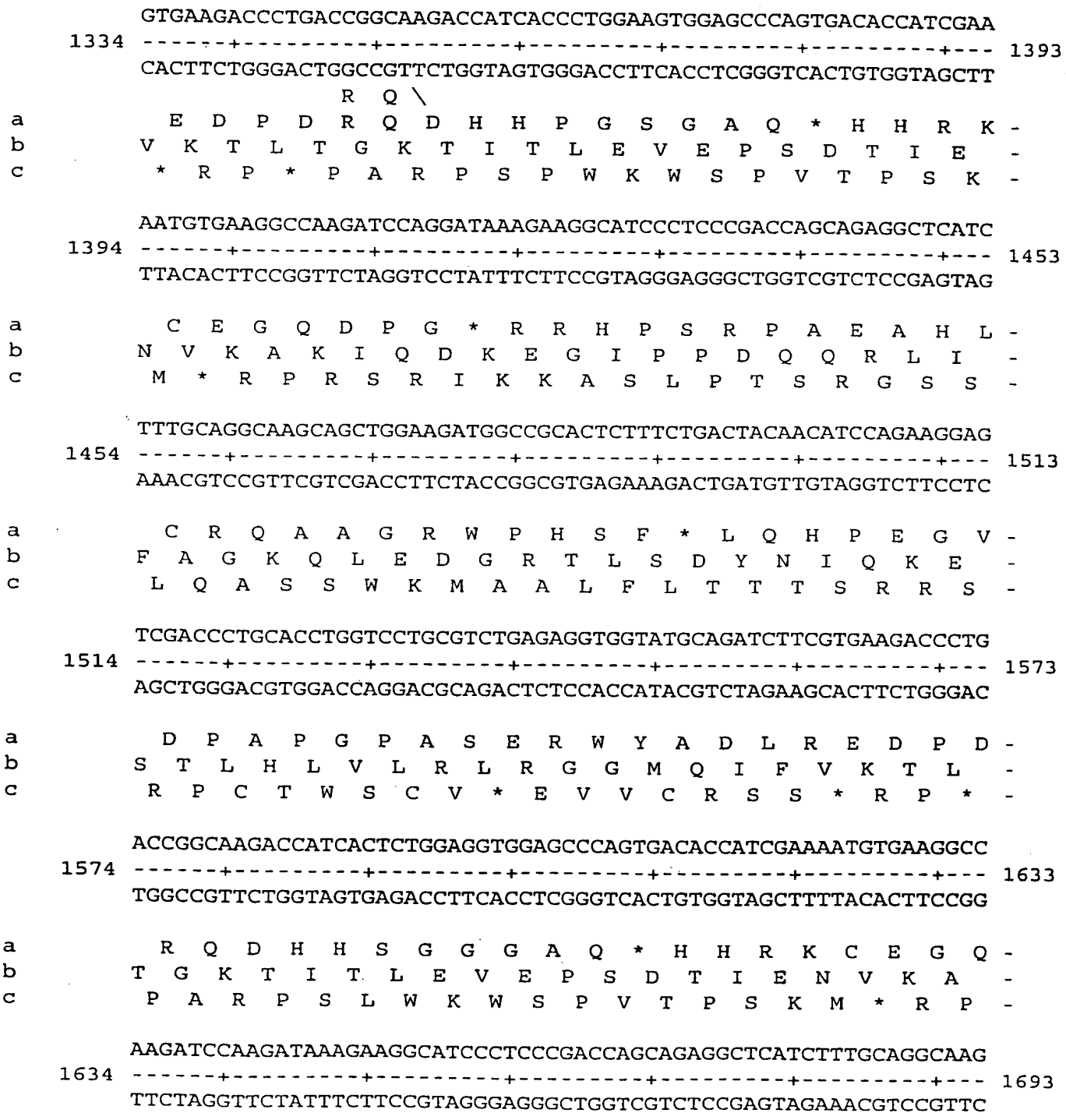
GTGAAGACCCTGACCGGCAAGACCATCACCCTGGAAGTGGAGCCCAGTGACACCATCGAA
1334 ------+---------+---------+---------+---------+---------+--- 1393
CACTTCTGGGACTGGCCGTTCTGGTAGTGGGACCTTCACCTCGGGTCACTGTGGTAGCTT
R Q \
a E D P D R Q D H H P G S G A Q * H H R K -
b V K T L T G K T I T L E V E P S D T I E -
c * R P * P A R P S P W K W S P V T P S K -
AATGTGAAGGCCAAGATCCAGGATAAAGAAGGCATCCCTCCCGACCAGCAGAGGCTCATC
1394 ------+---------+---------+---------+---------+---------+--- 1453
TTACACTTCCGGTTCTAGGTCCTATTTCTTCCGTAGGGAGGGCTGGTCGTCTCCGAGTAG
a C E G Q D P G * R R H P S R P A E A H L -
b N V K A K I Q D K E G I P P D Q Q R L I -
c M * R P R S R I K K A S L P T S R G S S -
TTTGCAGGCAAGCAGCTGGAAGATGGCCGCACTCTTTCTGACTACAACATCCAGAAGGAG
1454 ------+---------+---------+---------+---------+---------+--- 1513
AAACGTCCGTTCGTCGACCTTCTACCGGCGTGAGAAAGACTGATGTTGTAGGTCTTCCTC
a C R Q A A G R W P H S F * L Q H P E G V -
b F A G K Q L E D G R T L S D Y N I Q K E -
c L Q A S S W K M A A L F L T T T S R R S -
TCGACCCTGCACCTGGTCCTGCGTCTGAGAGGTGGTATGCAGATCTTCGTGAAGACCCTG
1514 ------+---------+---------+---------+---------+---------+--- 1573
AGCTGGGACGTGGACCAGGACGCAGACTCTCCACCATACGTCTAGAAGCACTTCTGGGAC
a D P A P G P A S E R W Y A D L R E D P D -
b S T L H L V L R L R G G M Q I F V K T L -
c R P C T W S C V * E V V C R S S * R P * -
ACCGGCAAGACCATCACTCTGGAGGTGGAGCCCAGTGACACCATCGAAAATGTGAAGGCC
1574 ------+---------+---------+---------+---------+---------+--- 1633
TGGCCGTTCTGGTAGTGAGACCTTCACCTCGGGTCACTGTGGTAGCTTTTACACTTCCGG
a R Q D H H S G G G A Q * H H R K C E G Q -
b T G K T I T L E V E P S D T I E N V K A -
c P A R P S L W K W S P V T P S K M * R P -
AAGATCCAAGATAAAGAAGGCATCCCTCCCGACCAGCAGAGGCTCATCTTTGCAGGCAAG
1634 ------+---------+---------+---------+---------+---------+--- 1693
TTCTAGGTTCTATTTCTTCCGTAGGGAGGGCTGGTCGTCTCCGAGTAGAAACGTCCGTTC

Figure 4-3

```
a          D  P  R  *  R  R  H  P  S  R  P  A  E  A  H  L  C  R  Q  A -
b        K  I  Q  D  K  E  G  I  P  P  D  Q  Q  R  L  I  F  A  G  K  -
c         R  S  K  I  K  K  A  S  L  P  T  S  R  G  S  S  L  Q  A  S -

         CAGCTGGAAGATGGCCGCACTCTTTCTGACTACAACATCCAGAAGGAGTCGACCCTGCAC
    1694 ------+---------+---------+---------+---------+---------+--- 1753
         GTCGACCTTCTACCGGCGTGAGAAAGACTGATGTTGTAGGTCTTCCTCAGCTGGGACGTG

a          A  G  R  W  P  H  S  F  *  L  Q  H  P  E  G  V  D  P  A  P -
b        Q  L  E  D  G  R  T  L  S  D  Y  N  I  Q  K  E  S  T  L  H  -
c         S  W  K  M  A  A  L  F  L  T  T  T  S  R  R  S  R  P  C  T -

         CTGGTCCTGCGCCTGAGGGGTGGCTGTTAATTCTTCAGTCATGGCAT
    1754 ------+---------+---------+---------+---------+ 1800
         GACCAGGACGCGGACTCCCCACCGACAATTAAGAAGTCAGTACCGTA

a          G  P  A  P  E  G  W  L  L  I  L  Q  S  W  H   -
b        L  V  L  R  L  R  G  G  C  *  F  F  S  H  G     -
c         W  S  C  A  *  G  V  A  V  N  S  S  V  M  A     -
```

Enzymes that do cut:

NONE

Enzymes that do not cut:

NotI

Figure 5-1 Apolipoprotein E

```
 (Linear) MAP of: hsapoe01.gcg  check: 2800  from: 1  to: 1157

ID   HSAPOE01   standard; RNA; HUM; 1157 BP.
XX
AC   M12529;
XX
NI   g178848
XX . . .

 With 1 enzymes: NOTI
                                October 31, 1996 15:09  ..

          ccccagcggaggtgaaggacgtccttccccaggagccgactggccaatcacaggcaggaa
        1 ---------+---------+---------+---------+---------+---------+ 60
          ggggtcgcctccacttcctgcaggaaggggtcctcggctgaccggttagtgtccgtcctt

a         P  Q  R  R  *  R  T  S  F  P  R  S  R  L  A  N  H  R  Q  E  -
b          P  S  G  G  E  G  R  P  S  P  G  A  D  W  P  I  T  G  R  K  -
c            P  A  E  V  K  D  V  L  P  Q  E  P  T  G  Q  S  Q  A  G  R -

          gatgaaggttctgtgggctgcgttgctggtcacattcctggcaggatgccaggccaaggt
       61 ---------+---------+---------+---------+---------+---------+ 120
          ctacttccaagacacccgacgcaacgaccagtgtaaggaccgtcctacggtccggttcca

a         D  E  G  S  V  G  C  V  A  G  H  I  P  G  R  M  P  G  Q  G  -
b          M  K  V  L  W  A  A  L  L  V  T  F  L  A  G  C  Q  A  K  V  -
c            *  R  F  C  G  L  R  C  W  S  H  S  W  Q  D  A  R  P  R  W -

          ggagcaagcggtggagacagagccggagcccgagctgcgccagcagaccgagtggcagag
      121 ---------+---------+---------+---------+---------+---------+ 180
          cctcgttcgccacctctgtctcggcctcgggctcgacgcggtcgtctggctcaccgtctc

a         G  A  S  G  G  D  R  A  G  A  R  A  A  P  A  D  R  V  A  E  -
b          E  Q  A  V  E  T  E  P  E  P  E  L  R  Q  Q  T  E  W  Q  S  -
c            S  K  R  W  R  Q  S  R  S  P  S  C  A  S  R  P  S  G  R  A -

          cggccagcgctgggaactggcactgggtcgcttttgggattacctgcgctgggtgcagac
      181 ---------+---------+---------+---------+---------+---------+ 240
          gccggtcgcgacccttgaccgtgacccagcgaaaaccctaatggacgcgacccacgtctg

a         R  P  A  L  G  T  G  T  G  S  L  L  G  L  P  A  L  G  A  D  -
b          G  Q  R  W  E  L  A  L  G  R  F  W  D  Y  L  R  W  V  Q  T  -
c            A  S  A  G  N  W  H  W  V  A  F  G  I  T  C  A  G  C  R  H -
```

Figure 5-2

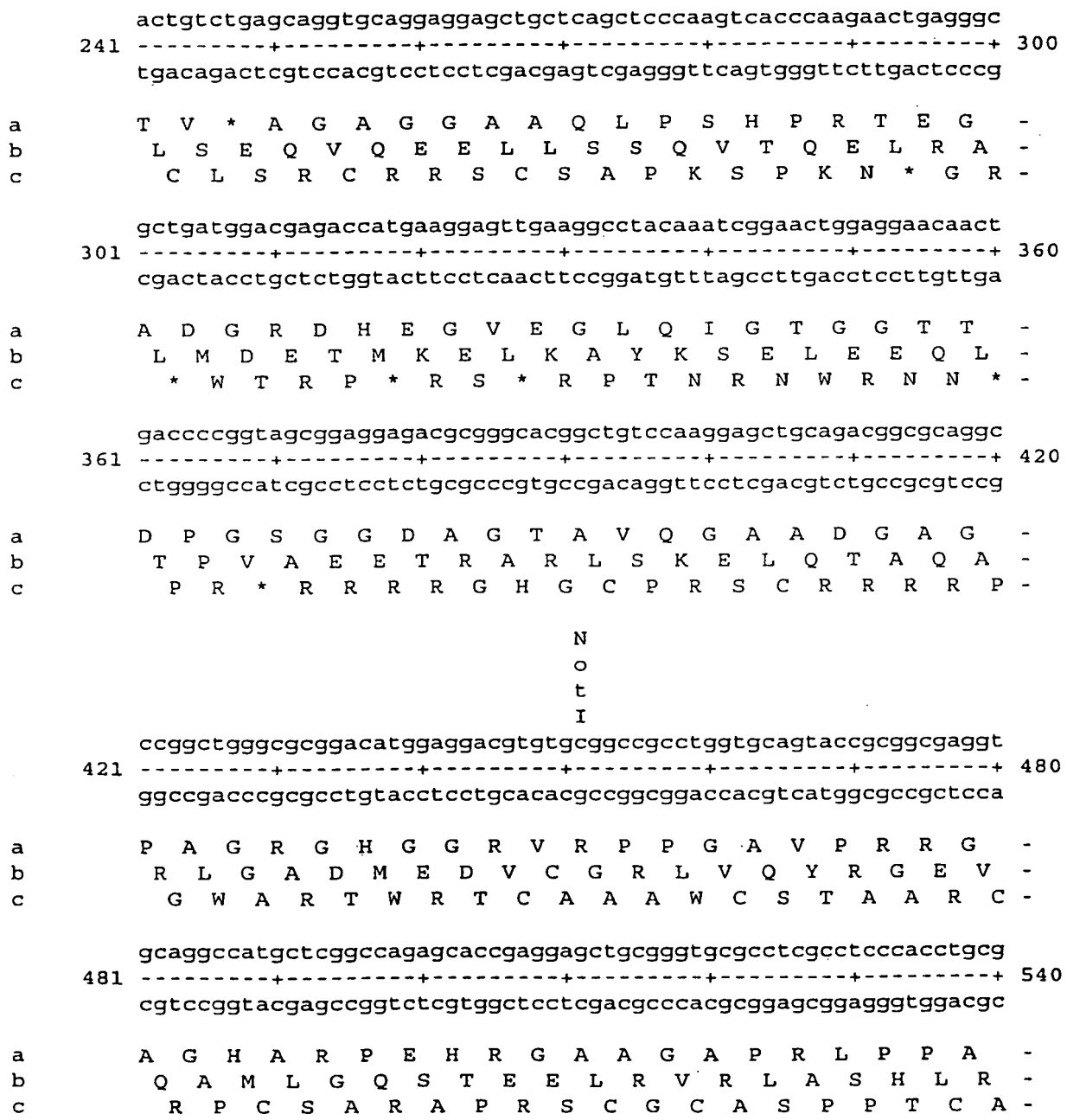
actgtctgagcaggtgcaggaggagctgctcagctcccaagtcacccaagaactgagggc
241
tgacagactcgtccacgtcctcctcgacgagtcgagggttcagtgggttcttgactcccg
a T V * A G A G G A A Q L P S H P R T E G -
b L S E Q V Q E E L L S S Q V T Q E L R A -
c C L S R C R R S C S A P K S P K N * G R -
gctgatggacgagaccatgaaggagttgaaggcctacaaatcggaactggaggaacaact
301
cgactacctgctctggtacttcctcaacttccggatgtttagccttgacctccttgttga
a A D G R D H E G V E G L Q I G T G G T T -
b L M D E T M K E L K A Y K S E L E E Q L -
c * W T R P * R S * R P T N R N W R N N * -
gaccccggtagcggaggagacgcgggcacggctgtccaaggagctgcagacggcgcaggc
361
ctggggccatcgcctcctctgcgcccgtgccgacaggttcctcgacgtctgccgcgtccg
a D P G S G G D A G T A V Q G A A D G A G -
b T P V A E E T R A R L S K E L Q T A Q A -
c P R * R R R R G H G C P R S C R R R R P -
N
o
t
I
ccggctgggcgcggacatggaggacgtgtgcggccgcctggtgcagtaccgcggcgaggt
421
ggccgacccgcgcctgtacctcctgcacacgccggcggaccacgtcatggcgccgctcca
a P A G R G H G G R V R P P G A V P R R G -
b R L G A D M E D V C G R L V Q Y R G E V -
c G W A R T W R T C A A A W C S T A A R C -
gcaggccatgctcggccagagcaccgaggagctgcgggtgcgcctcgcctcccacctgcg
481
cgtccggtacgagccggtctcgtggctcctcgacgcccacgcggagcggagggtggacgc
a A G H A R P E H R G A A G A P R L P P A -
b Q A M L G Q S T E E L R V R L A S H L R -
c R P C S A R A P R S C G C A S P P T C A -

Figure 5-3

```
        caagctgcgtaagcggctcctccgcgatcccgatgacctgcagaagcgcctggcagtgta
    541 ---------+---------+---------+---------+---------+---------+ 600
        gttcgacgcattcgccgaggaggcgctagggctactggacgtcttcgcggaccgtcacat

a       Q  A  A  *  A  A  P  P  R  S  R  *  P  A  E  A  P  G  S  V  -
b        K  L  R  K  R  L  L  R  D  P  D  D  L  Q  K  R  L  A  V  Y  -
c         S  C  V  S  G  S  S  A  I  P  M  T  C  R  S  A  W  Q  C  T -

        ccaggccggggcccgcgagggcgccgagcgcggcctcagcgccatccgcgagcgcctggg
    601 ---------+---------+---------+---------+---------+---------+ 660
        ggtccggccccgggcgctcccgcggctcgcgccggagtcgcggtaggcgctcgcggaccc

a       P  G  R  G  P  R  G  R  R  A  R  P  Q  R  H  P  R  A  P  G  -
b        Q  A  G  A  R  E  G  A  E  R  G  L  S  A  I  R  E  R  L  G  -
c         R  P  G  P  A  R  A  P  S  A  A  S  A  P  S  A  S  A  W  G -

        gcccctggtggaacagggccgcgtgcgggccgccactgtgggctccctggccggccagcc
    661 ---------+---------+---------+---------+---------+---------+ 720
        cggggaccaccttgtcccggcgcacgcccggcggtgacacccgagggaccggccggtcgg

a       A  P  G  G  T  G  P  R  A  G  R  H  C  G  L  P  G  R  P  A  -
b        P  L  V  E  Q  G  R  V  R  A  A  T  V  G  S  L  A  G  Q  P  -
c         P  W  W  N  R  A  A  C  G  P  P  L  W  A  P  W  P  A  S  R -

        gctacaggagcgggcccaggcctggggcgagcggctgcgcgcgcggatggaggagatggg
    721 ---------+---------+---------+---------+---------+---------+ 780
        cgatgtcctcgcccgggtccggaccccgctcgccgacgcgcgcgcctacctcctctaccc

a       A  T  G  A  G  P  G  L  G  R  A  A  A  R  A  D  G  G  D  G  -
b        L  Q  E  R  A  Q  A  W  G  E  R  L  R  A  R  M  E  E  M  G  -
c         Y  R  S  G  P  R  P  G  A  S  G  C  A  R  G  W  R  R  W  A -

        cagtcggacccgcgaccgcctggacgaggtgaaggagcaggtggcggaggtgcgcgccaa
    781 ---------+---------+---------+---------+---------+---------+ 840
        gtcagcctgggcgctggcggacctgctccacttcctcgtccaccgcctccacgcgcggtt

a       Q  S  D  P  R  P  P  G  R  G  E  G  A  G  G  G  G  A  R  Q  -
b        S  R  T  R  D  R  L  D  E  V  K  E  Q  V  A  E  V  R  A  K  -
c         V  G  P  A  T  A  W  T  R  *  R  S  R  W  R  R  C  A  P  S -

        gctggaggagcaggcccagcagatacgcctgcaggccgaggccttccaggcccgcctcaa
    841 ---------+---------+---------+---------+---------+---------+ 900
        cgacctcctcgtccgggtcgtctatgcggacgtccggctccggaaggtccgggcggagtt
```

Figure 5-4

```
a        A  G  G  A  G  P  A  D  T  P  A  G  R  G  L  P  G  P  P  Q  -
b         L  E  E  Q  A  Q  Q  I  R  L  Q  A  E  A  F  Q  A  R  L  K  -
c          W  R  S  R  P  S  R  Y  A  C  R  P  R  P  S  R  P  A  S  R  -

         gagctggttcgagcccctggtggaagacatgcagcgccagtgggccgggctggtggagaa
    901  ---------+---------+---------+---------+---------+---------+  960
         ctcgaccaagctcggggaccaccttctgtacgtcgcggtcacccggcccgaccacctctt

a        E  L  V  R  A  P  G  G  R  H  A  A  P  V  G  R  A  G  G  E  -
b         S  W  F  E  P  L  V  E  D  M  Q  R  Q  W  A  G  L  V  E  K  -
c          A  G  S  S  P  W  W  K  T  C  S  A  S  G  P  G  W  W  R  R  -

         ggtgcaggctgccgtgggcaccagcgccgccctgtgcccagcgacaatcactgaacgcc
    961  ---------+---------+---------+---------+---------+---------+ 1020
         ccacgtccgacggcacccgtggtcgcggcggggacacgggtcgctgttagtgacttgcgg

a        G  A  G  C  R  G  H  Q  R  R  P  C  A  Q  R  Q  S  L  N  A  -
b         V  Q  A  A  V  G  T  S  A  A  P  V  P  S  D  N  H  *  T  P  -
c          C  R  L  P  W  A  P  A  P  P  L  C  P  A  T  I  T  E  R  R  -

         gaagcctgcagccatgcgaccccacgccaccccgtgcctcctgcctccgcgcagcctgca
   1021  ---------+---------+---------+---------+---------+---------+ 1080
         cttcggacgtcggtacgctggggtgcggtggggcacggaggacggaggcgcgtcggacgt

a        E  A  C  S  H  A  T  P  R  H  P  V  P  P  A  S  A  Q  P  A  -
b         K  P  A  A  M  R  P  H  A  T  P  C  L  L  P  P  R  S  L  Q  -
c          S  L  Q  P  C  D  P  T  P  P  R  A  S  C  L  R  A  A  C  S  -

         gcgggagaccctgtccccgccccagccgtcctcctggggtggaccctagtttaataaaga
   1081  ---------+---------+---------+---------+---------+---------+ 1140
         cgccctctgggacaggggcggggtcggcaggaggaccccacctgggatcaaattatttct

a        A  G  D  P  V  P  A  P  A  V  L  L  G  W  T  L  V  *  *  R  -
b         R  E  T  L  S  P  P  Q  P  S  S  W  G  G  P  *  F  N  K  D  -
c          G  R  P  C  P  R  P  S  R  P  P  G  V  D  P  S  L  I  K  I  -

         ttcaccaagtttcacgc
   1141  ---------+-------  1157
         aagtggttcaaagtgcg

a        F  T  K  F  H     -
b         S  P  S  F  T     -
c          H  Q  V  S  R     -
```

Enzymes that do cut: NotI
Enzymes that do not cut: NONE

Figure 6-1 MAP2

(Linear) MAP of: hsmap2c.gcg check: 2201 from: 1 to: 5595

```
ID   HSMAP2C     standard; RNA; HUM; 5595 BP.
XX
AC   L12563;
XX
NI   g348216
XX . . .

With 1 enzymes: NOTI
                              October 31, 1996 14:28  ..

          atggctgacgagcggaaagacgaaggaaaggcacctcactggacctcagcaccgctaaca
        1 ---------+---------+---------+---------+---------+---------+ 60
          taccgactgctcgcctttctgcttcctttccgtggagtgacctggagtcgtggcgattgt

a         M  A  D  E  R  K  D  E  G  K  A  P  H  W  T  S  A  P  L  T  -
b          W  L  T  S  G  K  T  K  E  R  H  L  T  G  P  Q  H  R  *  Q  -
c           G  *  R  A  E  R  R  R  K  G  T  S  L  D  L  S  T  A  N  R -

          gaggcatctgcacactcacatccacctgagattaaggatcaaggcggagcaggggaagga
       61 ---------+---------+---------+---------+---------+---------+ 120
          ctccgtagacgtgtgagtgtaggtggactctaattcctagttccgcctcgtccccttcct

a         E  A  S  A  H  S  H  P  P  E  I  K  D  Q  G  G  A  G  E  G  -
b          R  H  L  H  T  H  I  H  L  R  L  R  I  K  A  E  Q  G  K  D  -
c           G  I  C  T  L  T  S  T  *  D  *  G  S  R  R  S  R  G  R  T -

          cttgtccgaagcgccaatggattcccatacagggaggatgaagagggtgcctttggagag
      121 ---------+---------+---------+---------+---------+---------+ 180
          gaacaggcttcgcggttacctaagggtatgtccctcctacttctcccacggaaacctctc

a         L  V  R  S  A  N  G  F  P  Y  R  E  D  E  E  G  A  F  G  E  -
b          L  S  E  A  P  M  D  S  H  T  G  R  M  K  R  V  P  L  E  S  -
c           C  P  K  R  Q  W  I  P  I  Q  G  G  *  R  G  C  L  W  R  A -

          catgggtcacagggcacctattcaaataccaaagagaatgggatcaacggagagctgacc
      181 ---------+---------+---------+---------+---------+---------+ 240
          gtacccagtgtcccgtggataagtttatggtttctcttaccctagttgcctctcgactgg

a         H  G  S  Q  G  T  Y  S  N  T  K  E  N  G  I  N  G  E  L  T  -
b          M  G  H  R  A  P  I  Q  I  P  K  R  M  G  S  T  E  S  *  P  -
c           W  V  T  G  H  L  F  K  Y  Q  R  E  W  D  Q  R  R  A  D  L -
```

Figure 6-2

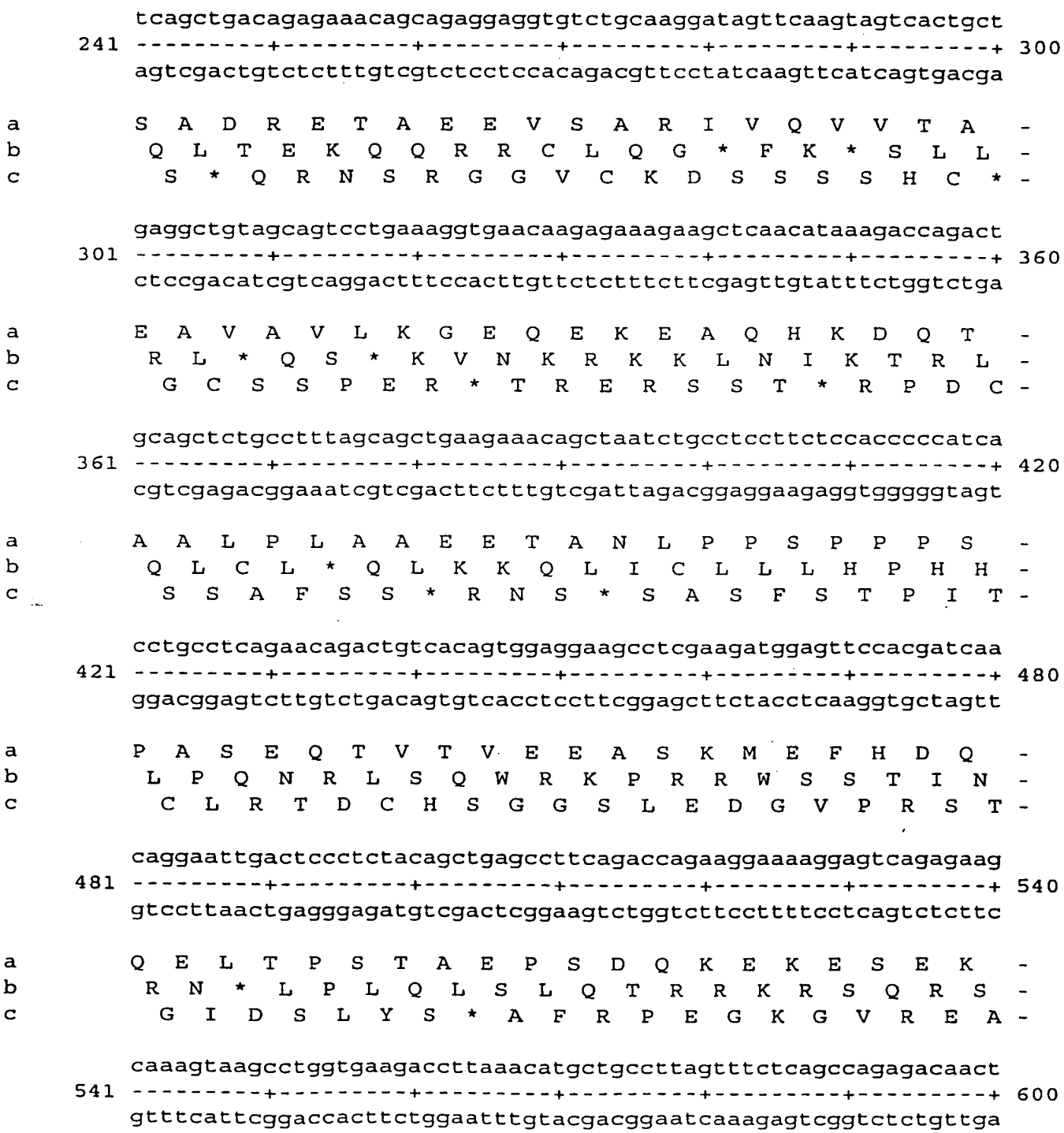
tcagctgacagagaaacagcagaggaggtgtctgcaaggatagttcaagtagtcactgct
241
agtcgactgtctctttgtcgtctcctccacagacgttcctatcaagttcatcagtgacga
300
a S A D R E T A E E V S A R I V Q V V T A -
b Q L T E K Q Q R R C L Q G * F K * S L L -
c S * Q R N S R G G V C K D S S S S H C * -
gaggctgtagcagtcctgaaaggtgaacaagagaaagaagctcaacataaagaccagact
301
ctccgacatcgtcaggactttccacttgttctctttcttcgagttgtatttctggtctga
360
a E A V A V L K G E Q E K E A Q H K D Q T -
b R L * Q S * K V N K R K K L N I K T R L -
c G C S S P E R * T R E R S S T * R P D C -
gcagctctgcctttagcagctgaagaaacagctaatctgcctccttctccacccccatca
361
cgtcgagacggaaatcgtcgacttctttgtcgattagacggaggaagaggtgggggtagt
420
a A A L P L A A E E T A N L P P S P P P S -
b Q L C L * Q L K K Q L I C L L L H P H H -
c S S A F S S * R N S * S A S F S T P I T -
cctgcctcagaacagactgtcacagtggaggaagcctcgaagatggagttccacgatcaa
421
ggacggagtcttgtctgacagtgtcacctccttcggagcttctacctcaaggtgctagtt
480
a P A S E Q T V T V E E A S K M E F H D Q -
b L P Q N R L S Q W R K P R R W S S T I N -
c C L R T D C H S G G S L E D G V P R S T -
caggaattgactccctctacagctgagccttcagaccagaaggaaaaggagtcagagaag
481
gtccttaactgagggagatgtcgactcggaagtctggtcttcctttcctcagtctcttc
540
a Q E L T P S T A E P S D Q K E K E S E K -
b R N * L P L Q L S L Q T R R K R S Q R S -
c G I D S L Y S * A F R P E G K G V R E A -
caaagtaagcctggtgaagaccttaaacatgctgccttagtttctcagccagagacaact
541
gtttcattcggaccacttctggaatttgtacgacggaatcaaagagtcggtctctgttga
600

Figure 6-3

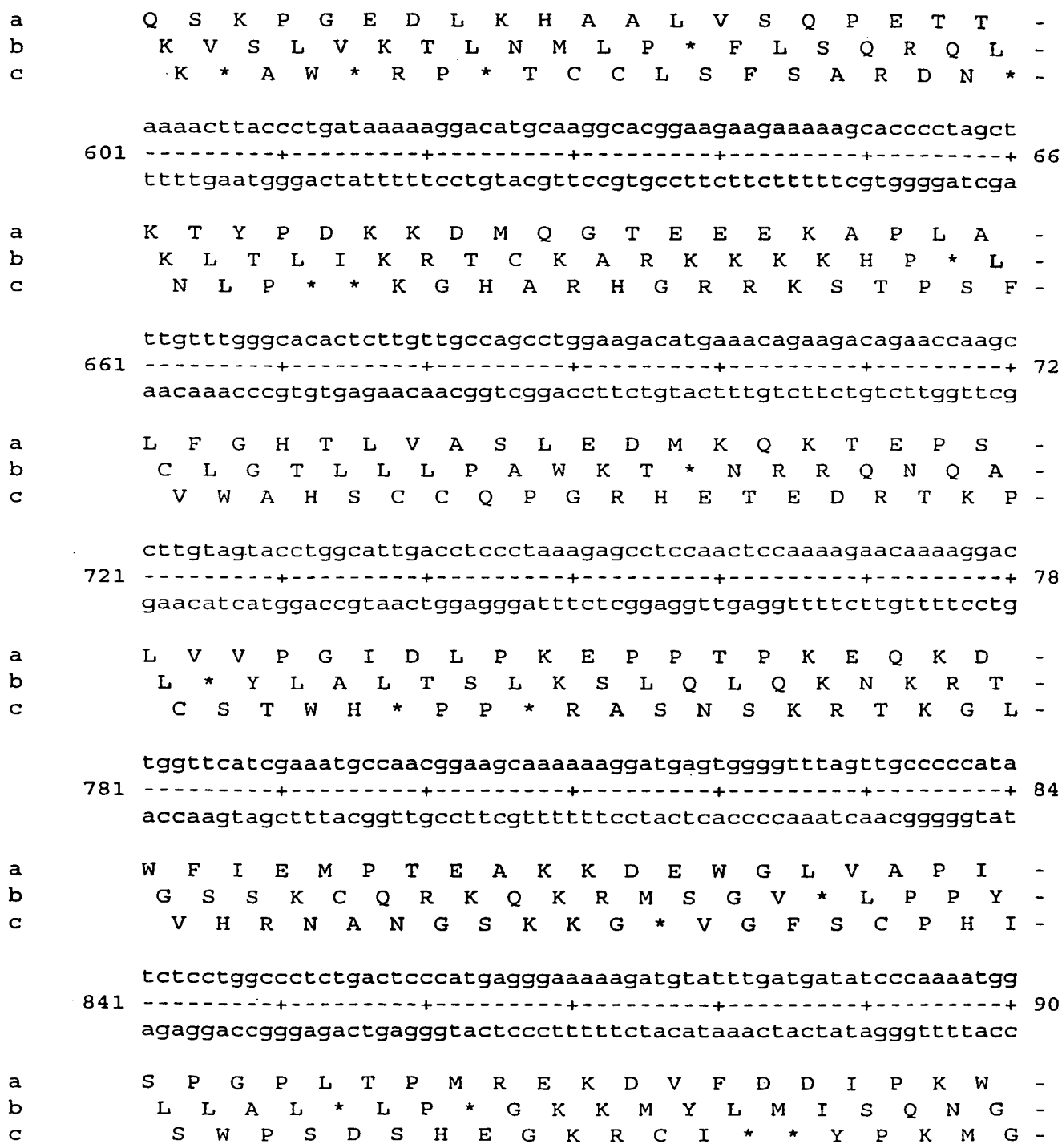
a Q S K P G E D L K H A A L V S Q P E T T -
b K V S L V K T L N M L P * F L S Q R Q L -
c K * A W * R P * T C C L S F S A R D N * -
aaaacttaccctgataaaaaggacatgcaaggcacggaagaagaaaaagcacccctagct
601 660
ttttgaatgggactatttttcctgtacgttccgtgccttcttcttttcgtggggatcga
a K T Y P D K K D M Q G T E E E K A P L A -
b K L T L I K R T C K A R K K K K H P * L -
c N L P * * K G H A R H G R R K S T P S F -
ttgtttgggcacactcttgttgccagcctggaagacatgaaacagaagacagaaccaagc
661 720
aacaaacccgtgtgagaacaacggtcggaccttctgtactttgtcttctgtcttggttcg
a L F G H T L V A S L E D M K Q K T E P S -
b C L G T L L L P A W K T * N R R Q N Q A -
c V W A H S C C Q P G R H E T E D R T K P -
cttgtagtacctggcattgacctccctaaagagcctccaactccaaaagaacaaaaggac
721 780
gaacatcatggaccgtaactggagggatttctcggaggttgaggttttcttgttttcctg
a L V V P G I D L P K E P P T P K E Q K D -
b L * Y L A L T S L K S L Q L Q K N K R T -
c C S T W H * P P * R A S N S K R T K G L -
tggttcatcgaaatgccaacggaagcaaaaaaggatgagtggggtttagttgcccccata
781 840
accaagtagctttacggttgccttcgtttttttcctactcaccccaaatcaacgggggtat
a W F I E M P T E A K K D E W G L V A P I -
b G S S K C Q R K Q K R M S G V * L P P Y -
c V H R N A N G S K K G * V G F S C P H I -
tctcctggccctctgactcccatgagggaaaaagatgtatttgatgatatcccaaaatgg
841 900
agaggaccgggagactgagggtactccctttttctacataaactactatagggttttacc
a S P G P L T P M R E K D V F D D I P K W -
b L L A L * L P * G K K M Y L M I S Q N G -
c S W P S D S H E G K R C I * * Y P K M G -

Figure 6-4

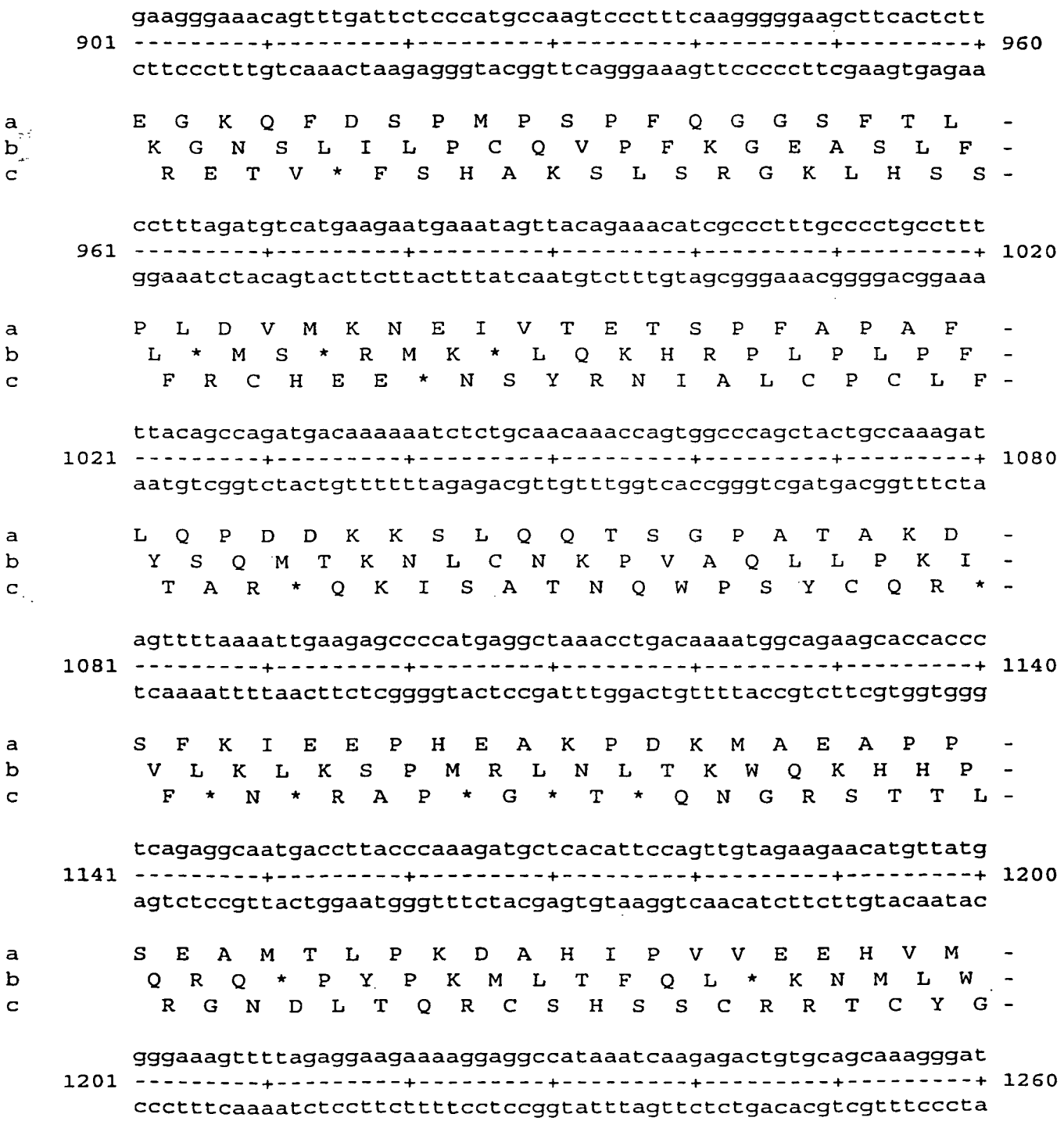
gaagggaaacagtttgattctcccatgccaagtccctttcaaggggaagcttcactctt
901 ---------+---------+---------+---------+---------+---------+ 960
cttccctttgtcaaactaagagggtacggttcagggaaagttccccccttcgaagtgagaa
a E G K Q F D S P M P S P F Q G G S F T L -
b K G N S L I L P C Q V P F K G E A S L F -
c R E T V * F S H A K S L S R G K L H S S -
cctttagatgtcatgaagaatgaaatagttacagaaacatcgccctttgcccctgccttt
961 ---------+---------+---------+---------+---------+---------+ 1020
ggaaatctacagtacttcttactttatcaatgtctttgtagcgggaaacggggacggaaa
a P L D V M K N E I V T E T S P F A P A F -
b L * M S * R M K * L Q K H R P L P L P F -
c F R C H E E * N S Y R N I A L C P C L F -
ttacagccagatgacaaaaaatctctgcaacaaaccagtggcccagctactgccaaagat
1021 ---------+---------+---------+---------+---------+---------+ 1080
aatgtcggtctactgtttttagagacgttgtttggtcaccgggtcgatgacggtttcta
a L Q P D D K K S L Q Q T S G P A T A K D -
b Y S Q M T K N L C N K P V A Q L L P K I -
c T A R * Q K I S A T N Q W P S Y C Q R * -
agttttaaaattgaagagccccatgaggctaaacctgacaaaatggcagaagcaccaccc
1081 ---------+---------+---------+---------+---------+---------+ 1140
tcaaaatttttaacttctcggggtactccgatttggactgttttaccgtcttcgtggtggg
a S F K I E E P H E A K P D K M A E A P P -
b V L K L K S P M R L N L T K W Q K H H P -
c F * N * R A P * G * T * Q N G R S T T L -
tcagaggcaatgaccttacccaaagatgctcacattccagttgtagaagaacatgttatg
1141 ---------+---------+---------+---------+---------+---------+ 1200
agtctccgttactggaatgggtttctacgagtgtaaggtcaacatcttcttgtacaatac
a S E A M T L P K D A H I P V V E E H V M -
b Q R Q * P Y P K M L T F Q L * K N M L W -
c R G N D L T Q R C S H S S C R R T C Y G -
gggaaagttttagaggaagaaaaggaggccataaatcaagagactgtgcagcaaagggat
1201 ---------+---------+---------+---------+---------+---------+ 1260
ccctttcaaaatctccttcttttcctccggtatttagttctctgacacgtcgtttcccta

Figure 6-5

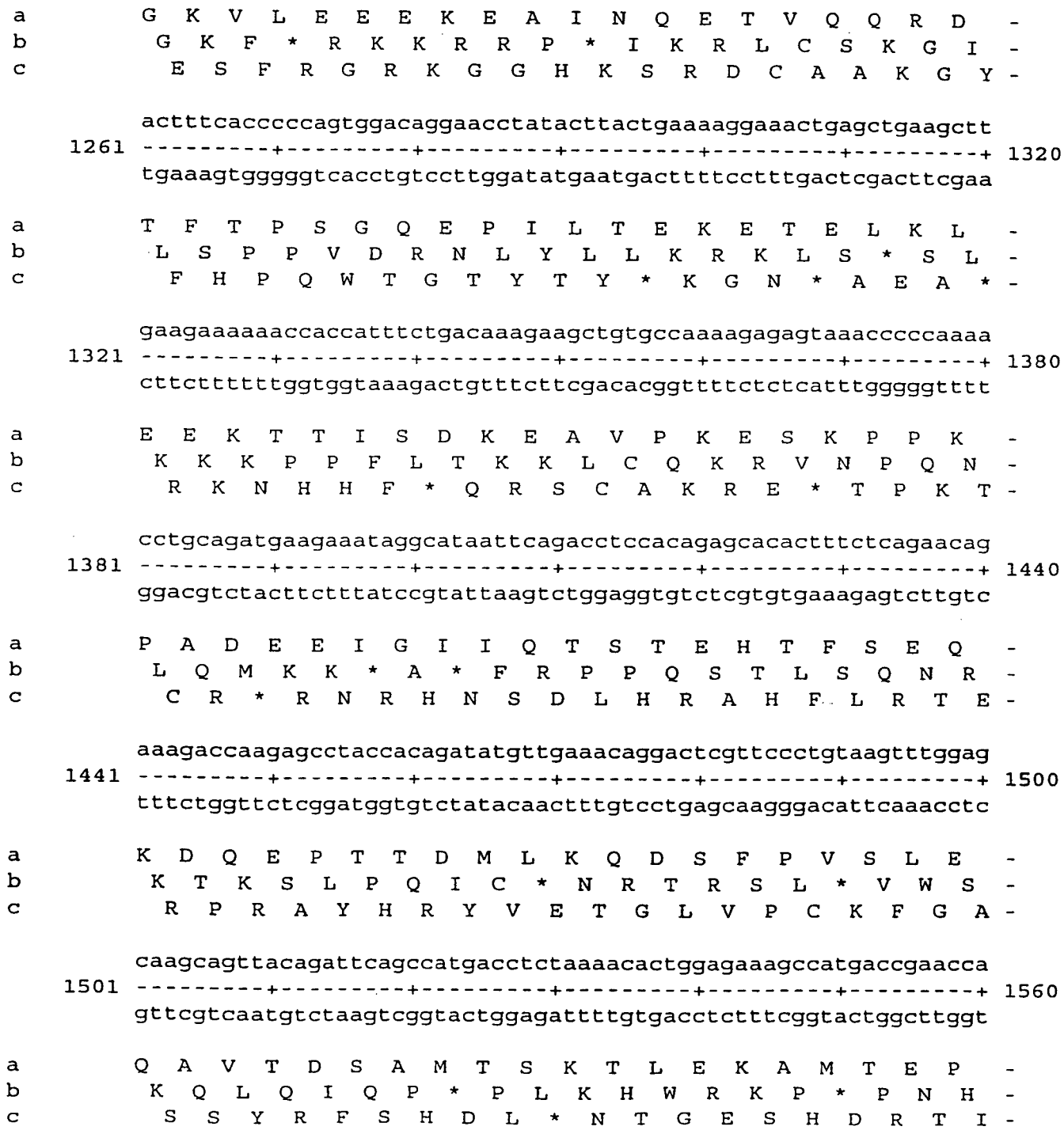
a G K V L E E E K E A I N Q E T V Q Q R D -
b G K F * R K K R R P * I K R L C S K G I -
c E S F R G R K G G H K S R D C A A K G Y -
actttcacccccagtggacaggaacctatacttactgaaaaggaaactgagctgaagctt
1261 ---------+---------+---------+---------+---------+---------+---------+ 1320
tgaaagtgggggtcacctgtccttggatatgaatgacttttcctttgactcgacttcgaa
a T F T P S G Q E P I L T E K E T E L K L -
b L S P P V D R N L Y L L K R K L S * S L -
c F H P Q W T G T Y T Y * K G N * A E A * -
gaagaaaaaaccaccatttctgacaaagaagctgtgccaaaagagagtaaacccccaaaa
1321 ---------+---------+---------+---------+---------+---------+---------+ 1380
cttctttttggtggtaaagactgtttcttcgacacggttttctctcatttgggggtttt
a E E K T T I S D K E A V P K E S K P P K -
b K K K P P F L T K K L C Q K R V N P Q N -
c R K N H H F * Q R S C A K R E * T P K T -
cctgcagatgaagaaataggcataattcagacctccacagagcacactttctcagaacag
1381 ---------+---------+---------+---------+---------+---------+---------+ 1440
ggacgtctacttctttatccgtattaagtctggaggtgtctcgtgtgaaagagtcttgtc
a P A D E E I G I I Q T S T E H T F S E Q -
b L Q M K K * A * F R P P Q S T L S Q N R -
c C R * R N R H N S D L H R A H F L R T E -
aaagaccaagagcctaccacagatatgttgaaacaggactcgttccctgtaagtttggag
1441 ---------+---------+---------+---------+---------+---------+---------+ 1500
tttctggttctcggatggtgtctatacaactttgtcctgagcaagggacattcaaacctc
a K D Q E P T T D M L K Q D S F P V S L E -
b K T K S L P Q I C * N R T R S L * V W S -
c R P R A Y H R Y V E T G L V P C K F G A -
caagcagttacagattcagccatgacctctaaaacactggagaaagccatgaccgaacca
1501 ---------+---------+---------+---------+---------+---------+---------+ 1560
gttcgtcaatgtctaagtcggtactggagattttgtgacctctttcggtactggcttggt
a Q A V T D S A M T S K T L E K A M T E P -
b K Q L Q I Q P * P L K H W R K P * P N H -
c S S Y R F S H D L * N T G E S H D R T I -

Figure 6-6

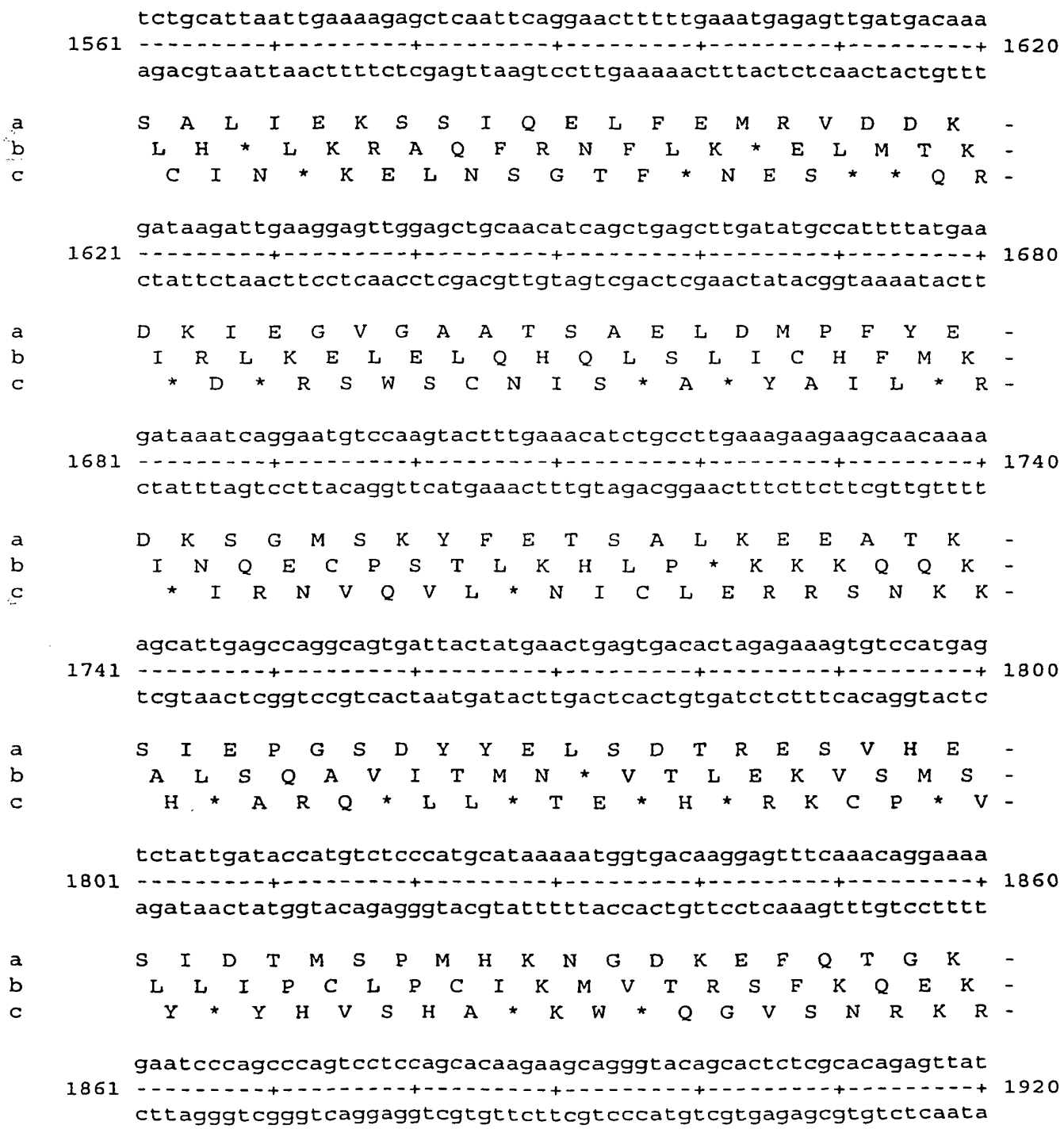
tctgcattaattgaaaagagctcaattcaggaacttttgaaatgagagttgatgacaaa
1561 ---------+---------+---------+---------+---------+---------+ 1620
agacgtaattaacttttctcgagttaagtccttgaaaaactttactctcaactactgttt
a S A L I E K S S I Q E L F E M R V D D K -
b L H * L K R A Q F R N F L K * E L M T K -
c C I N * K E L N S G T F * N E S * * Q R -
gataagattgaaggagttggagctgcaacatcagctgagcttgatatgccattttatgaa
1621 ---------+---------+---------+---------+---------+---------+ 1680
ctattctaacttcctcaacctcgacgttgtagtcgactcgaactatacggtaaaatactt
a D K I E G V G A A T S A E L D M P F Y E -
b I R L K E L E L Q H Q L S L I C H F M K -
c * D * R S W S C N I S * A * Y A I L * R -
gataaatcaggaatgtccaagtactttgaaacatctgccttgaaagaagaagcaacaaaa
1681 ---------+---------+---------+---------+---------+---------+ 1740
ctatttagtccttacaggttcatgaaactttgtagacggaactttcttcttcgttgtttt
a D K S G M S K Y F E T S A L K E E A T K -
b I N Q E C P S T L K H L P * K K K Q Q K -
c * I R N V Q V L * N I C L E R R S N K K -
agcattgagccaggcagtgattactatgaactgagtgacactagagaaagtgtccatgag
1741 ---------+---------+---------+---------+---------+---------+ 1800
tcgtaactcggtccgtcactaatgatacttgactcactgtgatctctttcacaggtactc
a S I E P G S D Y Y E L S D T R E S V H E -
b A L S Q A V I T M N * V T L E K V S M S -
c H * A R Q * L L * T E * H * R K C P * V -
tctattgataccatgtctcccatgcataaaaatggtgacaaggagtttcaaacaggaaaa
1801 ---------+---------+---------+---------+---------+---------+ 1860
agataactatggtacagagggtacgtatttttaccactgttcctcaaagtttgtcctttt
a S I D T M S P M H K N G D K E F Q T G K -
b L L I P C L P C I K M V T R S F K Q E K -
c Y * Y H V S H A * K W * Q G V S N R K R -
gaatcccagcccagtcctccagcacaagaagcagggtacagcactctcgcacagagttat
1861 ---------+---------+---------+---------+---------+---------+ 1920
cttagggtcgggtcaggaggtcgtgttcttcgtcccatgtcgtgagagcgtgtctcaata

Figure 6-7

```
a        E  S  Q  P  S  P  P  A  Q  E  A  G  Y  S  T  L  A  Q  S  Y  -
b         N  P  S  P  V  L  Q  H  K  K  Q  G  T  A  L  S  H  R  V  I  -
c          I  P  A  Q  S  S  S  T  R  S  R  V  Q  H  S  R  T  E  L  S -

         ccatcagatttacctgaagaacccagttctcctcaagaaagaatgttcactattgatcca
    1921 ---------+---------+---------+---------+---------+---------+---------+ 1980
         ggtagtctaaatggacttcttgggtcaagaggagttctttcttacaagtgataactaggt

a        P  S  D  L  P  E  E  P  S  S  P  Q  E  R  M  F  T  I  D  P  -
b         H  Q  I  Y  L  K  N  P  V  L  L  K  K  E  C  S  L  L  I  Q  -
c          I  R  F  T  *  R  T  Q  F  S  S  R  K  N  V  H  Y  *  S  K -

         aaagtgtatggagagaaaagggacctccacagtaagaataaggatgatttgacccttagc
    1981 ---------+---------+---------+---------+---------+---------+---------+ 2040
         tttcacatacctctcttttccctggaggtgtcattcttattcctactaaactgggaatcg

a        K  V  Y  G  E  K  R  D  L  H  S  K  N  K  D  D  L  T  L  S  -
b         K  C  M  E  R  K  G  T  S  T  V  R  I  R  M  I  *  P  L  A  -
c          S  V  W  R  E  K  G  P  P  Q  *  E  *  G  *  F  D  P  *  Q -

         aggagtttaggacttggtggtaggtctgcaatagaacaaagaagcatgtcaatcaatttg
    2041 ---------+---------+---------+---------+---------+---------+---------+ 2100
         tcctcaaatcctgaaccaccatccagacgttatcttgtttcttcgtacagttagttaaac

a        R  S  L  G  L  G  G  R  S  A  I  E  Q  R  S  M  S  I  N  L  -
b         G  V  *  D  L  V  V  G  L  Q  *  N  K  E  A  C  Q  S  I  C  -
c          E  F  R  T  W  W  *  V  C  N  R  T  K  K  H  V  N  Q  F  A -

         ccgatgtcttgcctagattccatagcccttggatttaactttggtcggggacatgatctt
    2101 ---------+---------+---------+---------+---------+---------+---------+ 2160
         ggctacagaacggatctaaggtatcgggaacctaaattgaaaccagcccctgtactagaa

a        P  M  S  C  L  D  S  I  A  L  G  F  N  F  G  R  G  H  D  L  -
b         R  C  L  A  *  I  P  *  P  L  D  L  T  L  V  G  D  M  I  F  -
c          D  V  L  P  R  F  H  S  P  W  I  *  L  W  S  G  T  *  S  F -

         tctcctctggcttccgatattctaaccaacactagtggaagtatggatgaaggggatgat
    2161 ---------+---------+---------+---------+---------+---------+---------+ 2220
         agaggagaccgaaggctataagattggttgtgatcaccttcatacctacttcccctacta

a        S  P  L  A  S  D  I  L  T  N  T  S  G  S  M  D  E  G  D  D  -
b         L  L  W  L  P  I  F  *  P  T  L  V  E  V  W  M  K  G  M  I  -
c          S  S  G  F  R  Y  S  N  Q  H  *  W  K  Y  G  *  R  G  *  L -
```

Figure 6-8

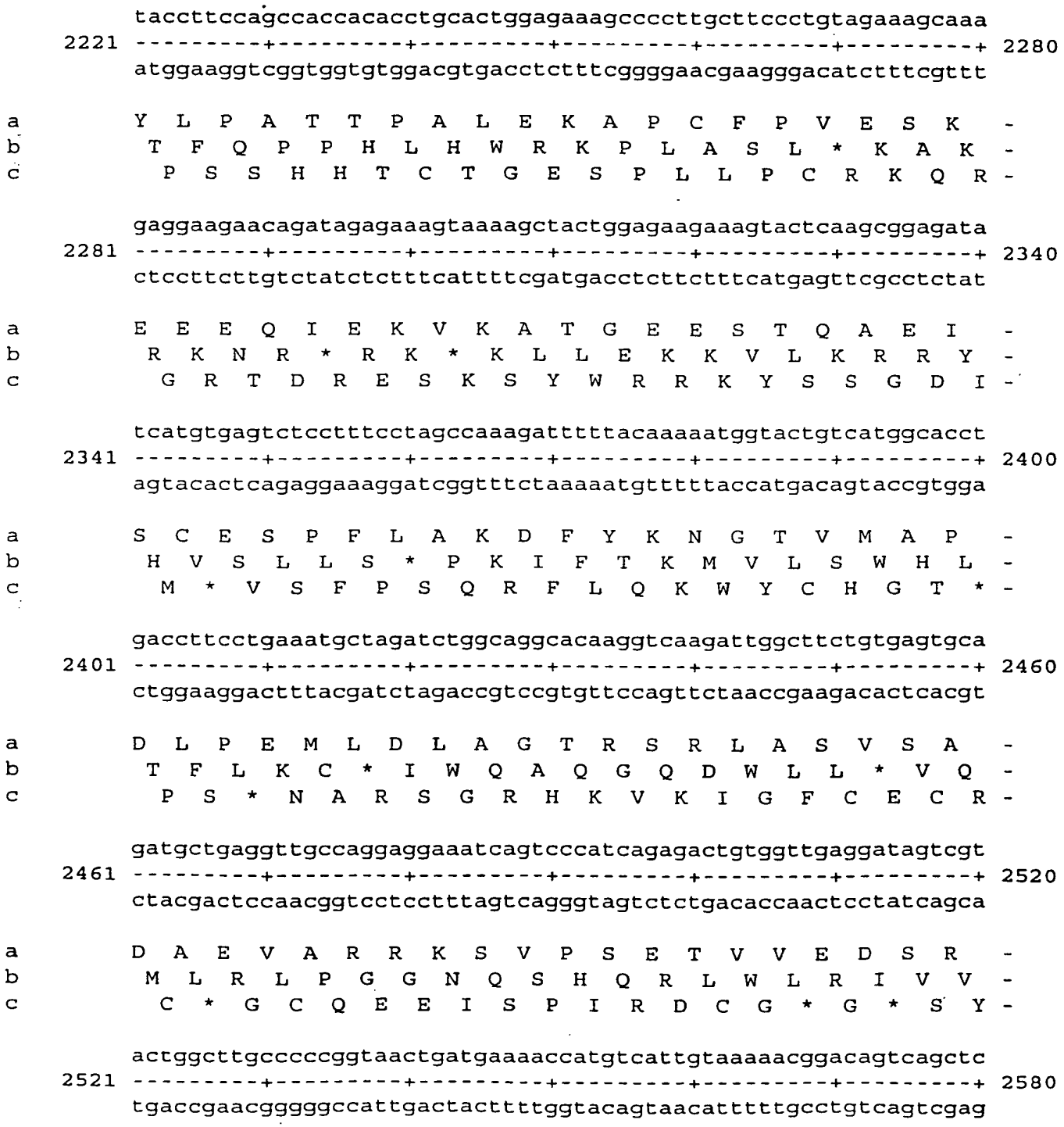
taccttccagccaccacacctgcactggagaaagccccttgcttccctgtagaaagcaaa
2221 ---------+---------+---------+---------+---------+---------+ 2280
atggaaggtcggtggtgtggacgtgacctctttcggggaacgaagggacatctttcgttt
a Y L P A T T P A L E K A P C F P V E S K -
b T F Q P P H L H W R K P L A S L * K A K -
c P S S H H T C T G E S P L L P C R K Q R -
gaggaagaacagatagagaaagtaaaagctactggagaagaaagtactcaagcggagata
2281 ---------+---------+---------+---------+---------+---------+ 2340
ctccttcttgtctatctctttcattttcgatgacctcttctttcatgagttcgcctctat
a E E E Q I E K V K A T G E E S T Q A E I -
b R K N R * R K * K L L E K K V L K R R Y -
c G R T D R E S K S Y W R R K Y S S G D I -
tcatgtgagtctcctttcctagccaaagattttacaaaaatggtactgtcatggcacct
2341 ---------+---------+---------+---------+---------+---------+ 2400
agtacactcagaggaaaggatcggtttctaaaaatgttttaccatgacagtaccgtgga
a S C E S P F L A K D F Y K N G T V M A P -
b H V S L L S * P K I F T K M V L S W H L -
c M * V S F P S Q R F L Q K W Y C H G T * -
gaccttcctgaaatgctagatctggcaggcacaaggtcaagattggcttctgtgagtgca
2401 ---------+---------+---------+---------+---------+---------+ 2460
ctggaaggactttacgatctagaccgtccgtgttccagttctaaccgaagacactcacgt
a D L P E M L D L A G T R S R L A S V S A -
b T F L K C * I W Q A Q G Q D W L L * V Q -
c P S * N A R S G R H K V K I G F C E C R -
gatgctgaggttgccaggaggaaatcagtcccatcagagactgtggttgaggatagtcgt
2461 ---------+---------+---------+---------+---------+---------+ 2520
ctacgactccaacggtcctcctttagtcagggtagtctctgacaccaactcctatcagca
a D A E V A R R K S V P S E T V V E D S R -
b M L R L P G G N Q S H Q R L W L R I V V -
c C * G C Q E E I S P I R D C G * G * S Y -
actggcttgcccccggtaactgatgaaaaccatgtcattgtaaaaacggacagtcagctc
2521 ---------+---------+---------+---------+---------+---------+ 2580
tgaccgaacgggggccattgactactttggtacagtaacattttgcctgtcagtcgag

Figure 6-9

```
a        T  G  L  P  P  V  T  D  E  N  H  V  I  V  K  T  D  S  Q  L  -
b         L  A  C  P  R  *  L  M  K  T  M  S  L  *  K  R  T  V  S  S  -
c          W  L  A  P  G  N  *  *  K  P  C  H  C  K  N  G  Q  S  A  R -

         gaagacctgggctactgtgtgttcaataagtacacagtcccattgccatcacctgttcaa
    2581 ---------+---------+---------+---------+---------+---------+---------+ 2640
         cttctggacccgatgacacacaagttattcatgtgtcagggtaacggtagtggacaagtt

a        E  D  L  G  Y  C  V  F  N  K  Y  T  V  P  L  P  S  P  V  Q  -
b         K  T  W  A  T  V  C  S  I  S  T  Q  S  H  C  H  H  L  F  K  -
c          R  P  G  L  L  C  V  Q  *  V  H  S  P  I  A  I  T  C  S  R -

         gacagtgagaatttatcaggggagagtggtacctttacgaaggcactgatgataaagtt
    2641 ---------+---------+---------+---------+---------+---------+---------+ 2700
         ctgtcactcttaaatagtcccctctcaccatggaaaatgcttccgtgactactatttcaa

a        D  S  E  N  L  S  G  E  S  G  T  F  Y  E  G  T  D  D  K  V  -
b         T  V  R  I  Y  Q  G  R  V  V  P  F  T  K  A  L  M  I  K  F  -
c          Q  *  E  F  I  R  G  E  W  Y  L  L  R  R  H  *  *  *  S  S -

         cgaagagatttggccacagacctttcactgattgaagtgaaactggcagcagccggaaga
    2701 ---------+---------+---------+---------+---------+---------+---------+ 2760
         gcttctctaaaccggtgtctggaaagtgactaacttcactttgaccgtcgtcggccttct

a        R  R  D  L  A  T  D  L  S  L  I  E  V  K  L  A  A  A  G  R  -
b         E  E  I  W  P  Q  T  F  H  *  L  K  *  N  W  Q  Q  P  E  E  -
c          K  R  F  G  H  R  P  F  T  D  *  S  E  T  G  S  S  R  K  S -

         gtcaaagatgagttcagtgttgacaaagaagcatccgcgcatatctctggtgacaaatca
    2761 ---------+---------+---------+---------+---------+---------+---------+ 2820
         cagtttctactcaagtcacaactgtttcttcgtaggcgcgtatagagaccactgtttagt

a        V  K  D  E  F  S  V  D  K  E  A  S  A  H  I  S  G  D  K  S  -
b         S  K  M  S  S  V  L  T  K  K  H  P  R  I  S  L  V  T  N  Q  -
c          Q  R  *  V  Q  C  *  Q  R  S  I  R  A  Y  L  W  *  Q  I  R -

         ggactgagtaaggagtttgaccaagagaagaaagctaatgataggttggatactgtacta
    2821 ---------+---------+---------+---------+---------+---------+---------+ 2880
         cctgactcattcctcaaactggttctcttctttcgattactatccaacctatgacatgat

a        G  L  S  K  E  F  D  Q  E  K  K  A  N  D  R  L  D  T  V  L  -
b         D  *  V  R  S  L  T  K  R  R  K  L  M  I  G  W  I  L  Y  *  -
c          T  E  *  G  V  *  P  R  E  E  S  *  *  *  V  G  Y  C  T  R -
```

Figure 6-10

```
          gaaaagagtgaagaacatgctgattcaaaagaacatgccaagaaaactgaagaggctggt
     2881 ---------+---------+---------+---------+---------+---------+ 2940
          cttttctcacttcttgtacgactaagttttcttgtacggttcttttgacttctccgacca

a         E  K  S  E  E  H  A  D  S  K  E  H  A  K  K  T  E  E  A  G  -
b          K  R  V  K  N  M  L  I  Q  K  N  M  P  R  K  L  K  R  L  V -
c           K  E  *  R  T  C  *  F  K  R  T  C  Q  E  N  *  R  G  W  * -

          gatgaaatagaaacattcggattaggagtaacctatgagcaagctttggccaaagatttg
     2941 ---------+---------+---------+---------+---------+---------+ 3000
          ctactttatctttgtaagcctaatcctcattggatactcgttcgaaaccggtttctaaac

a         D  E  I  E  T  F  G  L  G  V  T  Y  E  Q  A  L  A  K  D  L  -
b          M  K  *  K  H  S  D  *  E  *  P  M  S  K  L  W  P  K  I  C -
c           *  N  R  N  I  R  I  R  S  N  L  *  A  S  F  G  Q  R  F  V -

          tcaataccaacagatgcatcctctgagaaagcagagaagggtcttagttcagttccagag
     3001 ---------+---------+---------+---------+---------+---------+ 3060
          agttatggttgtctacgtaggagactctttcgtctcttcccagaatcaagtcaaggtctc

a         S  I  P  T  D  A  S  S  E  K  A  E  K  G  L  S  S  V  P  E  -
b          Q  Y  Q  Q  M  H  P  L  R  K  Q  R  R  V  L  V  Q  F  Q  R -
c           N  T  N  R  C  I  L  *  E  S  R  E  G  S  *  F  S  S  R  D -

          atagctgaggtagaaccatccaaaaaggtggaacaaggtctggattttgctgtccagggt
     3061 ---------+---------+---------+---------+---------+---------+ 3120
          tatcgactccatcttggtaggtttttccaccttgttccagacctaaaacgacaggtccca

a         I  A  E  V  E  P  S  K  K  V  E  Q  G  L  D  F  A  V  Q  G  -
b          *  L  R  *  N  H  P  K  R  W  N  K  V  W  I  L  L  S  R  V -
c           S  *  G  R  T  I  Q  K  G  G  T  R  S  G  F  C  C  P  G  S -

          caactagatgttaaaattagtgactttggacagatggcttcagggctaaacatagatgat
     3121 ---------+---------+---------+---------+---------+---------+ 3180
          gttgatctacaattttaatcactgaaacctgtctaccgaagtcccgatttgtatctacta

a         Q  L  D  V  K  I  S  D  F  G  Q  M  A  S  G  L  N  I  D  D  -
b          N  *  M  L  K  L  V  T  L  D  R  W  L  Q  G  *  T  *  M  I -
c           T  R  C  *  N  *  *  L  W  T  D  G  F  R  A  K  H  R  *  * -

          agaagggcaacagagctaaaacttgaggctacacaggacatgacccccctcatccaaagca
     3181 ---------+---------+---------+---------+---------+---------+ 3240
          tcttcccgttgtctcgattttgaactccgatgtgtcctgtactgggggagtaggtttcgt
```

Figure 6-11

```
a        R  R  A  T  E  L  K  L  E  A  T  Q  D  M  T  P  S  S  K  A  -
b         E  G  Q  Q  S  *  N  L  R  L  H  R  T  *  P  P  H  P  K  H  -
c          K  G  N  R  A  K  T  *  G  Y  T  G  H  D  P  L  I  Q  S  T -

         ccgcaggaggcagatgcatttatgggtgttgagtctggccacatgaaagaaggcactaaa
    3241 ---------+---------+---------+---------+---------+---------+ 3300
         ggcgtcctccgtctacgtaaatacccacaactcagaccggtgtactttcttccgtgattt

a        P  Q  E  A  D  A  F  M  G  V  E  S  G  H  M  K  E  G  T  K  -
b         R  R  R  Q  M  H  L  W  V  L  S  L  A  T  *  K  K  A  L  K  -
c          A  G  G  R  C  I  Y  G  C  *  V  W  P  H  E  R  R  H  *  S -

         gttagtgagacagaagtcaaacagaaggtggccaagcctgacttggtgcaccaggaggct
    3301 ---------+---------+---------+---------+---------+---------+ 3360
         caatcactctgtcttcagtttgtcttccaccggttcggactgaaccacgtggtcctccga

a        V  S  E  T  E  V  K  Q  K  V  A  K  P  D  L  V  H  Q  E  A  -
b         L  V  R  Q  K  S  N  R  R  W  P  S  L  T  W  C  T  R  R  L  -
c          *  *  D  R  S  Q  T  E  G  G  Q  A  *  L  G  A  P  G  G  C -

         gtagacaaggaggagtcctatgaatctagtggtgagcatgaaagtctcaccatggagtcc
    3361 ---------+---------+---------+---------+---------+---------+ 342(
         catctgttcctcctcaggatacttagatcaccactcgtactttcagagtggtacctcagg

a        V  D  K  E  E  S  Y  E  S  S  G  E  H  E  S  L  T  M  E  S  -
b         *  T  R  R  S  P  M  N  L  V  V  S  M  K  V  S  P  W  S  P  -
c          R  Q  G  G  V  L  *  I  *  W  *  A  *  K  S  H  H  G  V  L -

         ttgaaagctgatgagggcaagaaggaaacatctccagaatcatctctaattcaagatgag
    3421 ---------+---------+---------+---------+---------+---------+ 3480
         aactttcgactactcccgttcttcctttgtagaggtcttagtagagattaagttctactc

a        L  K  A  D  E  G  K  K  E  T  S  P  E  S  S  L  I  Q  D  E  -
b         *  K  L  M  R  A  R  R  K  H  L  Q  N  H  L  *  F  K  M  R  -
c          E  S  *  *  G  Q  E  G  N  I  S  R  I  I  S  N  S  R  *  D -

         attgccgtcaaattgtcagtggaaataccttgcccacctgctgtttcagaggctgattta
    3481 ---------+---------+---------+---------+---------+---------+ 3540
         taacggcagtttaacagtcacctttatggaacgggtggacgacaaagtctccgactaaat

a        I  A  V  K  L  S  V  E  I  P  C  P  P  A  V  S  E  A  D  L  -
b         L  P  S  N  C  Q  W  K  Y  L  A  H  L  L  F  Q  R  L  I  *  -
c          C  R  Q  I  V  S  G  N  T  L  P  T  C  C  F  R  G  *  F  S -
```

Figure 6-12

```
          gccacagatgagagagctgatgtccagatggaatttattcaggggccaaaagaagaaagc
     3541 ---------+---------+---------+---------+---------+---------+ 3600
          cggtgtctactctctcgactacaggtctaccttaaataagtccccggttttcttctttcg

a         A  T  D  E  R  A  D  V  Q  M  E  F  I  Q  G  P  K  E  E  S  -
b          P  Q  M  R  E  L  M  S  R  W  N  L  F  R  G  Q  K  K  K  A  -
c           H  R  *  E  S  *  C  P  D  G  I  Y  S  G  A  K  R  R  K  Q -

          aaagagaccccagatatatccatcacgccttctgatgttgcagagccattgcatgaaacg
     3601 ---------+---------+---------+---------+---------+---------+ 3660
          tttctctggggtctatataggtagtgcggaagactacaacgtctcggtaacgtactttgc

a         K  E  T  P  D  I  S  I  T  P  S  D  V  A  E  P  L  H  E  T  -
b          K  R  P  Q  I  Y  P  S  R  L  L  M  L  Q  S  H  C  M  K  R  -
c           R  D  P  R  Y  I  H  H  A  F  *  C  C  R  A  I  A  *  N  D -

          atcgtatctgaaccagcagagattcagagtgaggaagaagagatagaagcccagggagaa
     3661 ---------+---------+---------+---------+---------+---------+ 3720
          tagcatagacttggtcgtctctaagtctcactccttcttctctatcttcgggtccctctt

a         I  V  S  E  P  A  E  I  Q  S  E  E  E  E  I  E  A  Q  G  E  -
b          S  Y  L  N  Q  Q  R  F  R  V  R  K  K  R  *  K  P  R  E  N  -
c           R  I  *  T  S  R  D  S  E  *  G  R  R  D  R  S  P  G  R  I -

          tatgataaactgctcttccgctcagacacccttcagataactgacctgggtgtctcaggt
     3721 ---------+---------+---------+---------+---------+---------+ 3780
          atactatttgacgagaaggcgagtctgtgggaagtctattgactggacccacagagtcca

a         Y  D  K  L  L  F  R  S  D  T  L  Q  I  T  D  L  G  V  S  G  -
b          M  I  N  C  S  S  A  Q  T  P  F  R  *  L  T  W  V  S  Q  V  -
c           *  *  T  A  L  P  L  R  H  P  S  D  N  *  P  G  C  L  R  C -

          gccagggaggaatttgtggagacctgcccaagtgaacacaaaggagtgattgagtctgtt
     3781 ---------+---------+---------+---------+---------+---------+ 3840
          cggtccctccttaaacacctctggacgggttcacttgtgtttcctcactaactcagacaa

a         A  R  E  E  F  V  E  T  C  P  S  E  H  K  G  V  I  E  S  V  -
b          P  G  R  N  L  W  R  P  A  Q  V  N  T  K  E  *  L  S  L  L  -
c           Q  G  G  I  C  G  D  L  P  K  *  T  Q  R  S  D  *  V  C  C -

          gtgaccatcgaggatgatttcatcactgtagtgcaaaccacaactgatgaaggggagtca
     3841 ---------+---------+---------+---------+---------+---------+ 3900
          cactggtagctcctactaaagtagtgacatcacgtttggtgttgactacttcccctcagt
```

Figure 6-13

```
a        V  T  I  E  D  D  F  I  T  V  V  Q  T  T  T  D  E  G  E  S  -
b         *  P  S  R  M  I  S  S  L  *  C  K  P  Q  L  M  K  G  S  Q  -
c          D  H  R  G  *  F  H  H  C  S  A  N  H  N  *  *  R  G  V  R -

         gggtcccacagcgtgcgttttgcagccctagagcagcctgaggtggaaaggagaccatct
    3901 ---------+---------+---------+---------+---------+---------+ 3960
         cccagggtgtcgcacgcaaaacgtcgggatctcgtcggactccacctttcctctggtaga

a        G  S  H  S  V  R  F  A  A  L  E  Q  P  E  V  E  R  R  P  S  -
b         G  P  T  A  C  V  L  Q  P  *  S  S  L  R  W  K  G  D  H  L  -
c          V  P  Q  R  A  F  C  S  P  R  A  A  *  G  G  K  E  T  I  S -

         cctcatgatgaagaagagtttgaagtagaagaggcagctgaagcccaggcagaacccaaa
    3961 ---------+---------+---------+---------+---------+---------+ 4020
         ggagtactacttcttctcaaacttcatcttctccgtcgacttcgggtccgtcttgggttt

a        P  H  D  E  E  E  F  E  V  E  E  A  A  E  A  Q  A  E  P  K  -
b         L  M  M  K  K  S  L  K  *  K  R  Q  L  K  P  R  Q  N  P  K  -
c          S  *  *  R  R  V  *  S  R  R  G  S  *  S  P  G  R  T  Q  R -

         gatggttccccagaggctccagcttcccctgagagagaagaggttgcactttctgaatat
    4021 ---------+---------+---------+---------+---------+---------+ 408(
         ctaccaaggggtctccgaggtcgaaggggactctctcttctccaacgtgaaagacttata

a        D  G  S  P  E  A  P  A  S  P  E  R  E  E  V  A  L  S  E  Y  -
b         M  V  P  Q  R  L  Q  L  P  L  R  E  K  R  L  H  F  L  N  I  -
c          W  F  P  R  G  S  S  F  P  *  E  R  R  G  C  T  F  *  I  * -

         aagacagaaacctatgacgattacaaagatgagaccaccattgacgactccatcatggac
    4081 ---------+---------+---------+---------+---------+---------+ 4140
         ttctgtctttggatactgctaatgtttctactctggtggtaactgctgaggtagtacctg

a        K  T  E  T  Y  D  D  Y  K  D  E  T  T  I  D  D  S  I  M  D  -
b         R  Q  K  P  M  T  I  T  K  M  R  P  P  L  T  T  P  S  W  T  -
c          D  R  N  L  *  R  L  Q  R  *  D  H  H  *  R  L  H  H  G  R -

         gctgacagcctctgggtggacactcaagatgatgataggagcatcatgacagaacagtta
    4141 ---------+---------+---------+---------+---------+---------+ 4200
         cgactgtcggagacccacctgtgagttctactactatcctcgtagtactgtcttgtcaat

a        A  D  S  L  W  V  D  T  Q  D  D  D  R  S  I  M  T  E  Q  L  -
b         L  T  A  S  G  W  T  L  K  M  M  I  G  A  S  *  Q  N  S  *  -
c          *  Q  P  L  G  G  H  S  R  *  *  *  E  H  H  D  R  T  V  R -
```

Figure 6-14

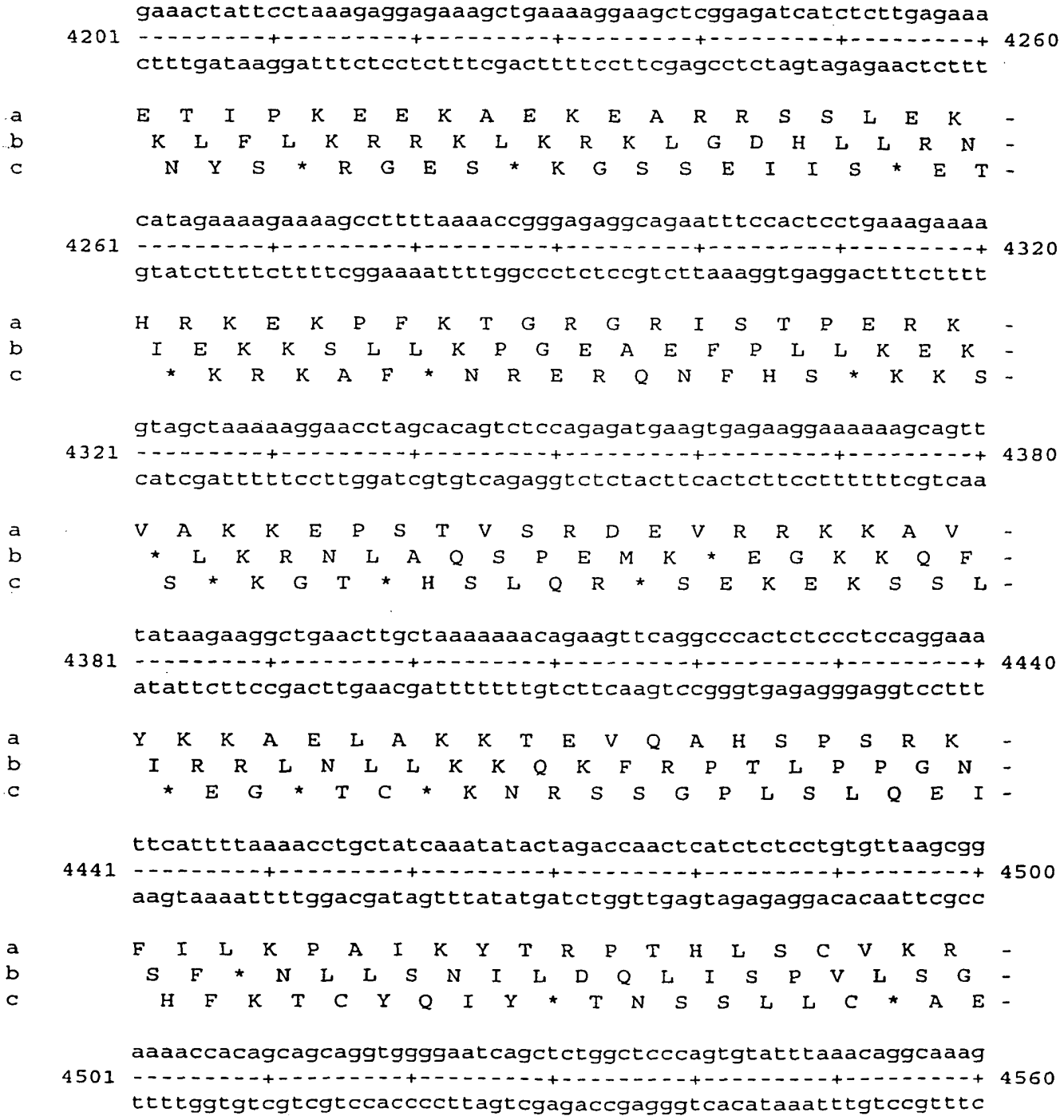
gaaactattcctaaagaggagaaagctgaaaaggaagctcggagatcatctcttgagaaa
4201 ---------+---------+---------+---------+---------+---------+ 4260
ctttgataaggatttctcctctttcgacttttccttcgagcctctagtagagaactcttt
a E T I P K E E K A E K E A R R S S L E K -
b K L F L K R R K L K R K L G D H L L R N -
c N Y S * R G E S * K G S S E I I S * E T -
catagaaaagaaaagccttttaaaaccgggagaggcagaatttccactcctgaaagaaaa
4261 ---------+---------+---------+---------+---------+---------+ 4320
gtatcttttcttttcggaaaattttggccctctccgtcttaaaggtgaggactttctttt
a H R K E K P F K T G R G R I S T P E R K -
b I E K K S L L K P G E A E F P L L K E K -
c * K R K A F * N R E R Q N F H S * K K S -
gtagctaaaaaggaacctagcacagtctccagagatgaagtgagaaggaaaaaagcagtt
4321 ---------+---------+---------+---------+---------+---------+ 4380
catcgatttttccttggatcgtgtcagaggtctctacttcactcttccttttttcgtcaa
a V A K K E P S T V S R D E V R R K K A V -
b * L K R N L A Q S P E M K * E G K K Q F -
c S * K G T * H S L Q R * S E K E K S S L -
tataagaaggctgaacttgctaaaaaaacagaagttcaggcccactctccctccaggaaa
4381 ---------+---------+---------+---------+---------+---------+ 4440
atattcttccgacttgaacgatttttttgtcttcaagtccgggtgagagggaggtccttt
a Y K K A E L A K K T E V Q A H S P S R K -
b I R R L N L L K K Q K F R P T L P P G N -
c * E G * T C * K N R S S G P L S L Q E I -
ttcattttaaaacctgctatcaaatatactagaccaactcatctctcctgtgttaagcgg
4441 ---------+---------+---------+---------+---------+---------+ 4500
aagtaaaattttggacgatagtttatatgatctggttgagtagagaggacacaattcgcc
a F I L K P A I K Y T R P T H L S C V K R -
b S F * N L L S N I L D Q L I S P V L S G -
c H F K T C Y Q I Y * T N S S L L C * A E -
aaaaccacagcagcaggtggggaatcagctctggctcccagtgtatttaaacaggcaaag
4501 ---------+---------+---------+---------+---------+---------+ 4560
ttttggtgtcgtcgtccacccttagtcgagaccgagggtcacataaatttgtccgtttc

Figure 6-15

```
a        K  T  T  A  A  G  G  E  S  A  L  A  P  S  V  F  K  Q  A  K  -
b         K  P  Q  Q  Q  V  G  N  Q  L  W  L  P  V  Y  L  N  R  Q  R  -
c          N  H  S  S  R  W  G  I  S  S  G  S  Q  C  I  *  T  G  K  G -

         gacaaagtctctgacggagtaaccaagagcccagaaaagcgctcttctctcccaagacct
    4561 ---------+---------+---------+---------+---------+---------+ 4620
         ctgtttcagagactgcctcattggttctcgggtcttttcgcgagaagagagggttctgga

a        D  K  V  S  D  G  V  T  K  S  P  E  K  R  S  S  L  P  R  P  -
b         T  K  S  L  T  E  *  P  R  A  Q  K  S  A  L  L  S  Q  D  L  -
c          Q  S  L  *  R  S  N  Q  E  P  R  K  A  L  F  S  P  K  T  F -

         tcctccattctccctcctcggcgaggtgtgtcaggagacagagatgagaattccttctct
    4621 ---------+---------+---------+---------+---------+---------+ 4680
         aggaggtaagagggaggagccgctccacacagtcctctgtctctactcttaaggaagaga

a        S  S  I  L  P  P  R  R  G  V  S  G  D  R  D  E  N  S  F  S  -
b         P  P  F  S  L  L  G  E  V  C  Q  E  T  E  M  R  I  P  S  L  -
c          L  H  S  P  S  S  A  R  C  V  R  R  Q  R  *  E  F  L  L  S -

         ctcaacagttctatctcttcttcagcacggcggaccaccaggtcagagccaattcgcaga
    4681 ---------+---------+---------+---------+---------+---------+ 4740
         gagttgtcaagatagagaagaagtcgtgccgcctggtggtccagtctcggttaagcgtct

a        L  N  S  S  I  S  S  S  A  R  R  T  T  R  S  E  P  I  R  R  -
b         S  T  V  L  S  L  L  Q  H  G  G  P  P  G  Q  S  Q  F  A  E  -
c          Q  Q  F  Y  L  F  F  S  T  A  D  H  Q  V  R  A  N  S  Q  S -

         gcagggaagagtggtacctcaacacccactacccctgggtctactgccatcactcctggc
    4741 ---------+---------+---------+---------+---------+---------+ 4800
         cgtcccttctcaccatggagttgtgggtgatggggacccagatgacggtagtgaggaccg

a        A  G  K  S  G  T  S  T  P  T  T  P  G  S  T  A  I  T  P  G  -
b         Q  G  R  V  V  P  Q  H  P  L  P  L  G  L  L  P  S  L  L  A  -
c          R  E  E  W  Y  L  N  T  H  Y  P  W  V  Y  C  H  H  S  W  H -

         accccaccaagttattcttcacgcacaccaggcactcctggaacccctagctatcccagg
    4801 ---------+---------+---------+---------+---------+---------+ 4860
         tggggtggttcaataagaagtgcgtgtggtccgtgaggaccttggggatcgatagggtcc

a        T  P  P  S  Y  S  S  R  T  P  G  T  P  G  T  P  S  Y  P  R  -
b         P  H  Q  V  I  L  H  A  H  Q  A  L  L  E  P  L  A  I  P  G  -
c          P  T  K  L  F  F  T  H  T  R  H  S  W  N  P  *  L  S  Q  D -
```

Figure 6-16

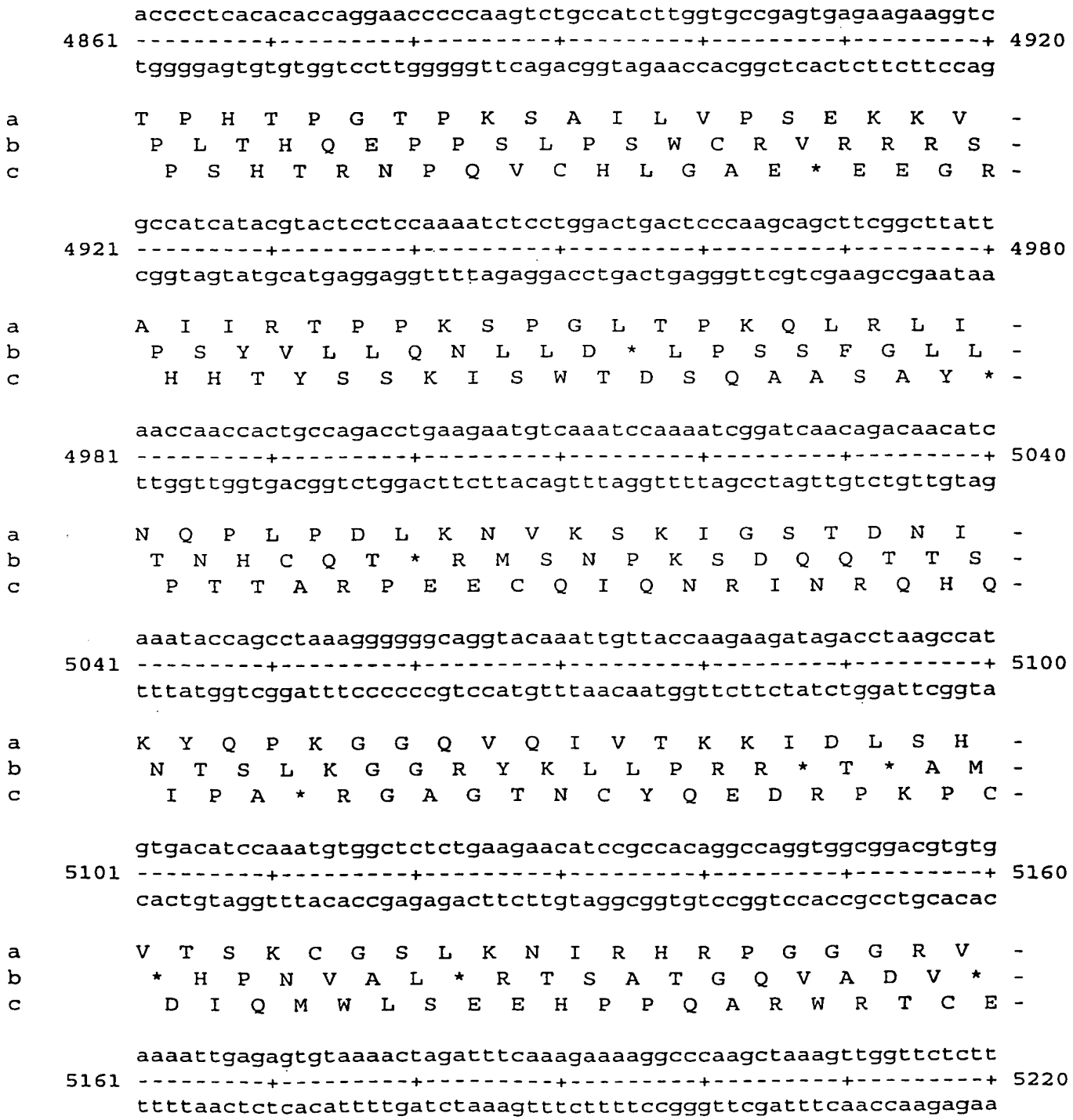
acccctcacacaccaggaaccccaagtctgccatcttggtgccgagtgagaagaaggtc
4861 ---------+---------+---------+---------+---------+---------+ 4920
tggggagtgtgtggtccttgggggttcagacggtagaaccacggctcactcttcttccag
a T P H T P G T P K S A I L V P S E K K V -
b P L T H Q E P P S L P S W C R V R R R S -
c P S H T R N P Q V C H L G A E * E E G R -
gccatcatacgtactcctccaaaatctcctggactgactcccaagcagcttcggcttatt
4921 ---------+---------+---------+---------+---------+---------+ 4980
cggtagtatgcatgaggaggttttagaggacctgactgagggttcgtcgaagccgaataa
a A I I R T P P K S P G L T P K Q L R L I -
b P S Y V L L Q N L L D * L P S S F G L L -
c H H T Y S S K I S W T D S Q A A S A Y * -
aaccaaccactgccagacctgaagaatgtcaaatccaaaatcggatcaacagacaacatc
4981 ---------+---------+---------+---------+---------+---------+ 5040
ttggttggtgacggtctggacttcttacagtttaggttttagcctagttgtctgttgtag
a N Q P L P D L K N V K S K I G S T D N I -
b T N H C Q T * R M S N P K S D Q Q T T S -
c P T T A R P E E C Q I Q N R I N R Q H Q -
aaataccagcctaaaggggggcaggtacaaattgttaccaagaagatagacctaagccat
5041 ---------+---------+---------+---------+---------+---------+ 5100
tttatggtcggatttccccccgtccatgtttaacaatggttcttctatctggattcggta
a K Y Q P K G G Q V Q I V T K K I D L S H -
b N T S L K G G R Y K L L P R R * T * A M -
c I P A * R G A G T N C Y Q E D R P K P C -
gtgacatccaaatgtggctctctgaagaacatccgccacaggccaggtggcggacgtgtg
5101 ---------+---------+---------+---------+---------+---------+ 5160
cactgtaggtttacaccgagagacttcttgtaggcggtgtccggtccaccgcctgcacac
a V T S K C G S L K N I R H R P G G G R V -
b * H P N V A L * R T S A T G Q V A D V * -
c D I Q M W L S E E H P P Q A R W R T C E -
aaaattgagagtgtaaaactagatttcaaagaaaaggcccaagctaaagttggttctctt
5161 ---------+---------+---------+---------+---------+---------+ 5220
ttttaactctcacattttgatctaaagtttcttttccgggttcgatttcaaccaagagaa

Figure 6-17

```
a        K  I  E  S  V  K  L  D  F  K  E  K  A  Q  A  K  V  G  S  L  -
b         K  L  R  V  *  N  *  I  S  K  K  R  P  K  L  K  L  V  L  L  -
c          N  *  E  C  K  T  R  F  Q  R  K  G  P  S  *  S  W  F  S  *  -

         gataatgctcatcatgtacctggaggtggtaatgtcaagattgacagccaaaagttgaac
    5221 ---------+---------+---------+---------+---------+---------+---------+ 5280
         ctattacgagtagtacatggacctccaccattacagttctaactgtcggttttcaacttg

a        D  N  A  H  H  V  P  G  G  G  N  V  K  I  D  S  Q  K  L  N  -
b         I  M  L  I  M  Y  L  E  V  V  M  S  R  L  T  A  K  S  *  T  -
c          *  C  S  S  C  T  W  R  W  *  C  Q  D  *  Q  P  K  V  E  L  -

         ttcagagagcatgctaaagcccgtgtggaccatggggctgagatcattacacagtcccca
    5281 ---------+---------+---------+---------+---------+---------+---------+ 5340
         aagtctctcgtacgatttcgggcacacctggtaccccgactctagtaatgtgtcaggggt

a        F  R  E  H  A  K  A  R  V  D  H  G  A  E  I  I  T  Q  S  P  -
b         S  E  S  M  L  K  P  V  W  T  M  G  L  R  S  L  H  S  P  Q  -
c          Q  R  A  C  *  S  P  C  G  P  W  G  *  D  H  Y  T  V  P  R  -

         ggcagatccagcgtggcatcaccccgacgactcagcaatgtctcctcgtctggaagcatc
    5341 ---------+---------+---------+---------+---------+---------+---------+ 540C
         ccgtctaggtcgcaccgtagtggggctgctgagtcgttacagaggagcagaccttcgtag

a        G  R  S  S  V  A  S  P  R  R  L  S  N  V  S  S  S  G  S  I  -
b         A  D  P  A  W  H  H  P  D  D  S  A  M  S  P  R  L  E  A  S  -
c          Q  I  Q  R  G  I  T  P  T  T  Q  Q  C  L  L  V  W  K  H  Q  -

         aacctgctcgaatctcctcagcttgccactttggctgaggatgtcactgctgcactcgct
    5401 ---------+---------+---------+---------+---------+---------+---------+ 5460
         ttggacgagcttagaggagtcgaacggtgaaaccgactcctacagtgacgacgtgagcga

a        N  L  L  E  S  P  Q  L  A  T  L  A  E  D  V  T  A  A  L  A  -
b         T  C  S  N  L  L  S  L  P  L  W  L  R  M  S  L  L  H  S  L  -
c          P  A  R  I  S  S  A  C  H  F  G  *  G  C  H  C  C  T  R  *  -

         aagcagggcttgtgaatatttctcatttagcattgaaataataatatttaggcatgagct
    5461 ---------+---------+---------+---------+---------+---------+---------+ 5520
         ttcgtcccgaacacttataaagagtaaatcgtaactttattattataaatccgtactcga

a        K  Q  G  L  *  I  F  L  I  *  H  *  N  N  N  I  *  A  *  A  -
b         S  R  A  C  E  Y  F  S  F  S  I  E  I  I  I  F  R  H  E  L  -
c          A  G  L  V  N  I  S  H  L  A  L  K  *  *  Y  L  G  M  S  S  -
```

Figure 6-18

```
          cttggcaggagtgggctctgagcagttgttatatcattctttataaaccataaaataaat
     5521 ---------+---------+---------+---------+---------+---------+ 5580
          gaaccgtcctcacccgagactcgtcaacaatatagtaagaaatatttggtattttattta

a           L  G  R  S  G  L  *  A  V  V  I  S  F  F  I  N  H  K  I  N  -
b            L  A  G  V  G  S  E  Q  L  L  Y  H  S  L  *  T  I  K  *  I  -
c             W  Q  E  W  A  L  S  S  C  Y  I  I  L  Y  K  P  *  N  K  *  -

          aatctcccggaattc
     5581 ---------+----- 5595
          ttagagggccttaag

a           N  L  P  E  F  -
b            I  S  R  N     -
c             S  P  G  I     -
```

Enzymes that do cut:

NONE

Enzymes that do not cut:

NotI

Figure 7-1 Neurofilament L

(Linear) MAP of: hsnflg.gcg check: 5926 from: 1 to: 4682

ID HSNFLG standard; DNA; HUM; 4682 BP.
XX
AC X05608; S42443;
XX
NI e1002618
XX . . .

With 1 enzymes: NOTI

October 31, 1996 14:33 ..

```
          aaggatccaagtgtcacggggtctgggcaatgcaggacgggaggggctgcgtgagtgagt
        1 ---------+---------+---------+---------+---------+---------+ 60
          ttcctaggttcacagtgccccagacccgttacgtcctgccctccccgacgcactcactca

a           K  D  P  S  V  T  G  S  G  Q  C  R  T  G  G  A  A  *  V  S  -
b            R  I  Q  V  S  R  G  L  G  N  A  G  R  E  G  L  R  E  *  V  -
c             G  S  K  C  H  G  V  W  A  M  Q  D  G  R  G  C  V  S  E  Y -

          acagaagggaaatgagtgaggggcatgggatctcagagaaaatcagggacctctgagca
       61 ---------+---------+---------+---------+---------+---------+ 120
          tgtcttccctttactcactccccgtaccctagagtctcttttagtccctggagactcgt

a           T  E  G  K  *  V  R  G  H  G  I  S  E  K  I  R  D  L  *  A  -
b            Q  K  G  N  E  *  G  G  M  G  S  Q  R  K  S  G  T  S  E  Q  -
c             R  R  E  M  S  E  G  A  W  D  L  R  E  N  Q  G  P  L  S  K -

          aagtggaaaggacgaccgccgcagctcctcgggccgtagctcgaccccgccttccctttt
      121 ---------+---------+---------+---------+---------+---------+ 180
          ttcacctttcctgctggcggcgtcgaggagcccggcatcgagctggggcggaagggaaaa

a           K  W  K  G  R  P  P  Q  L  L  G  P  *  L  D  P  A  F  P  F  -
b            S  G  K  D  D  R  R  S  S  S  G  R  S  S  T  P  P  S  L  F  -
c             V  E  R  T  T  A  A  A  P  R  A  V  A  R  P  R  L  P  F  S -

          ccgcagaatcctcgccttggctgcagcagcgcgctgccccactggccggcgtgccgtga
      181 ---------+---------+---------+---------+---------+---------+ 240
          ggcgtcttaggagcggaaccgacgtcgtcgcgcgacgggggtgaccggccgcacggcact

a           P  Q  N  P  R  L  G  C  S  S  A  L  P  P  L  A  G  V  P  *  -
b            R  R  I  L  A  L  A  A  A  A  R  C  P  H  W  P  A  C  R  D  -
c             A  E  S  S  P  W  L  Q  Q  R  A  A  P  T  G  R  R  A  V  I -
```

Figure 7-2

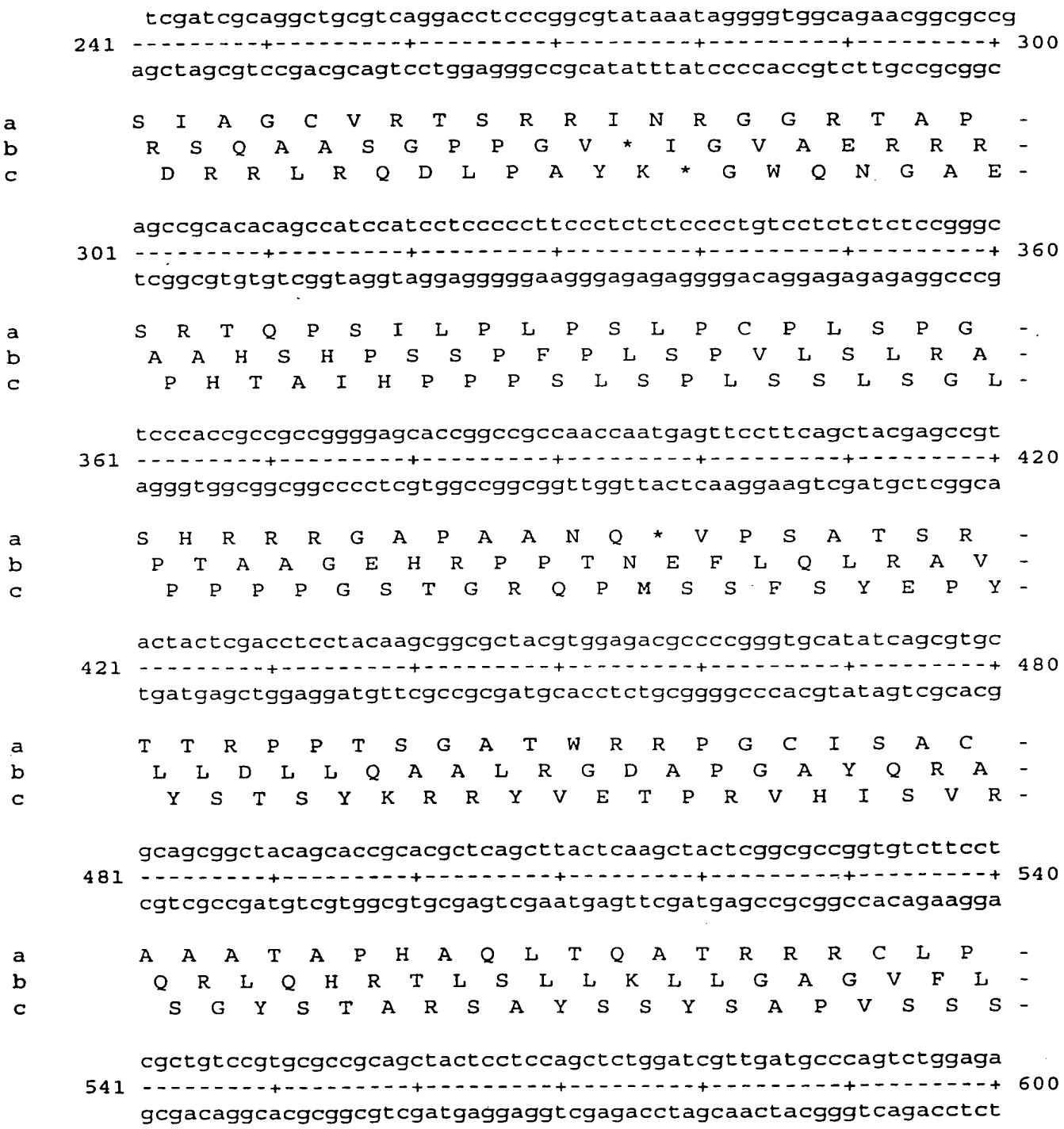
tcgatcgcaggctgcgtcaggacctcccggcgtataaataggggtggcagaacggcgccg
241 ---------+---------+---------+---------+---------+---------+ 300
agctagcgtccgacgcagtcctggagggccgcatatttatccccaccgtcttgccgcggc
a S I A G C V R T S R R I N R G G R T A P -
b R S Q A A S G P P G V * I G V A E R R R -
c D R R L R Q D L P A Y K * G W Q N G A E -
agccgcacacagccatccatcctccccttccctctctccccctgtcctctctctccgggc
301 ---------+---------+---------+---------+---------+---------+ 360
tcggcgtgtgtcggtaggtaggaggggaagggagagaggggacaggagagaggcccg
a S R T Q P S I L P L P S L P C P L S P G -
b A A H S H P S S P F P L S P V L S L R A -
c P H T A I H P P P S L S P L S S L S G L -
tcccaccgccgccggggagcaccggccgccaaccaatgagttccttcagctacgagccgt
361 ---------+---------+---------+---------+---------+---------+ 420
agggtggcggcggcccctcgtggccggcggttggttactcaaggaagtcgatgctcggca
a S H R R R G A P A A N Q * V P S A T S R -
b P T A A G E H R P P T N E F L Q L R A V -
c P P P P G S T G R Q P M S S F S Y E P Y -
actactcgacctcctacaagcggcgctacgtggagacgccccgggtgcatatcagcgtgc
421 ---------+---------+---------+---------+---------+---------+ 480
tgatgagctggaggatgttcgccgcgatgcacctctgcggggcccacgtatagtcgcacg
a T T R P P T S G A T W R R P G C I S A C -
b L L D L L Q A A L R G D A P G A Y Q R A -
c Y S T S Y K R R Y V E T P R V H I S V R -
gcagcggctacagcaccgcacgctcagcttactcaagctactcggcgccggtgtcttcct
481 ---------+---------+---------+---------+---------+---------+ 540
cgtcgccgatgtcgtggcgtgcgagtcgaatgagttcgatgagccgcggccacagaagga
a A A A T A P H A Q L T Q A T R R R C L P -
b Q R L Q H R T L S L L K L L G A G V F L -
c S G Y S T A R S A Y S S Y S A P V S S S -
cgctgtccgtgcgccgcagctactcctccagctctggatcgttgatgcccagtctggaga
541 ---------+---------+---------+---------+---------+---------+ 600
gcgacaggcacgcggcgtcgatgaggaggtcgagacctagcaactacgggtcagacctct

Figure 7-3

```
a       R  C  P  C  A  A  A  T  P  P  A  L  D  R  *  C  P  V  W  R  -
b        A  V  R  A  P  Q  L  L  L  Q  L  W  I  V  D  A  Q  S  G  E  -
c         L  S  V  R  R  S  Y  S  S  S  S  G  S  L  M  P  S  L  E  N -

        acctcgacctgagccaggtagccgccatcagcaacgacctcaagtccatccgcacgcagg
    601 ---------+---------+---------+---------+---------+---------+---------+ 660
        tggagctggactcggtccatcggcggtagtcgttgctggagttcaggtaggcgtgcgtcc

a       T  S  T  *  A  R  *  P  P  S  A  T  T  S  S  P  S  A  R  R  -
b        P  R  P  E  P  G  S  R  H  Q  Q  R  P  Q  V  H  P  H  A  G  -
c         L  D  L  S  Q  V  A  A  I  S  N  D  L  K  S  I  R  T  Q  E -

        agaaggcgcagctccaggacctcaatgaccgcttcgccagcttcatcgagcgcgtgcacg
    661 ---------+---------+---------+---------+---------+---------+---------+ 720
        tcttccgcgtcgaggtcctggagttactggcgaagcggtcgaagtagctcgcgcacgtgc

a       R  R  R  S  S  R  T  S  M  T  A  S  P  A  S  S  S  A  C  T  -
b        E  G  A  A  P  G  P  Q  *  P  L  R  Q  L  H  R  A  R  A  R  -
c         K  A  Q  L  Q  D  L  N  D  R  F  A  S  F  I  E  R  V  H  E -

        agctggagcagcagaacaaggtcctggaagccgagctgctggtgctgcgccagaagcact
    721 ---------+---------+---------+---------+---------+---------+---------+ 780
        tcgacctcgtcgtcttgttccaggaccttcggctcgacgaccacgacgcggtcttcgtga

a       S  W  S  S  R  T  R  S  W  K  P  S  C  W  C  C  A  R  S  T  -
b        A  G  A  A  E  Q  G  P  G  S  R  A  A  G  A  A  P  E  A  L  -
c         L  E  Q  Q  N  K  V  L  E  A  E  L  L  V  L  R  Q  K  H  S -

        ccgagccatcccgcttccgggcgctgtacgagcaggagatccgcgacctgcgcctagcgg
    781 ---------+---------+---------+---------+---------+---------+---------+ 840
        ggctcggtagggcgaaggcccgcgacatgctcgtcctctaggcgctggacgcggatcgcc

a       P  S  H  P  A  S  G  R  C  T  S  R  R  S  A  T  C  A  *  R  -
b        R  A  I  P  L  P  G  A  V  R  A  G  D  P  R  P  A  P  S  G  -
c         E  P  S  R  F  R  A  L  Y  E  Q  E  I  R  D  L  R  L  A  A -

        cggaagatgccaccaccaacgagaagcaagcgctccgaggcgagcgcgaagaagggctgg
    841 ---------+---------+---------+---------+---------+---------+---------+ 900
        gccttctacggtggtggttgctcttcgttcgcgaggctccgctcgcgcttcttcccgacc

a       R  K  M  P  P  P  T  R  S  K  R  S  E  A  S  A  K  K  G  W  -
b        G  R  C  H  H  Q  R  E  A  S  A  P  R  R  A  R  R  R  A  G  -
c         E  D  A  T  T  N  E  K  Q  A  L  R  G  E  R  E  E  G  L  E -
```

Figure 7-4

```
        aggagaccctgcgcaacctgcaggcgcgctatgaagaggaggtgctgagccgcgaggacg
    901 ---------+---------+---------+---------+---------+---------+ 960
        tcctctgggacgcgttggacgtccgcgcgatacttctcctccacgactcggcgctcctgc

a       R  R  P  C  A  T  C  R  R  A  M  K  R  R  C  *  A  A  R  T   -
b        G  D  P  A  Q  P  A  G  A  L  *  R  G  G  A  E  P  R  G  R  -
c         E  T  L  R  N  L  Q  A  R  Y  E  E  E  V  L  S  R  E  D  A -

        ccgagggccggctgatggaacgccgcaaaggcgccgacgaggcggcgctcgctcgcgccg
    961 ---------+---------+---------+---------+---------+---------+ 1020
        ggctcccggccgactaccttgcggcgtttccgcggctgctccgccgcgagcgagcgcggc

a       P  R  A  G  *  W  N  A  A  K  A  P  T  R  R  R  S  L  A  P   -
b        R  G  P  A  D  G  T  P  Q  R  R  R  R  G  G  A  R  S  R  R  -
c         E  G  R  L  M  E  R  R  K  G  A  D  E  A  A  L  A  R  A  E -

        agctcgagaagcgcatcgacagcttgatggacgaaatctcttttctgaagaaagtgcacg
   1021 ---------+---------+---------+---------+---------+---------+ 1080
        tcgagctcttcgcgtagctgtcgaactacctgctttagagaaaagacttctttcacgtgc

a       S  S  R  S  A  S  T  A  *  W  T  K  S  L  F  *  R  K  C  T   -
b        A  R  E  A  H  R  Q  L  D  G  R  N  L  F  S  E  E  S  A  R  -
c         L  E  K  R  I  D  S  L  M  D  E  I  S  F  L  K  K  V  H  E -

        aagaggagatcgccgaactgcaggcgcagatccagtacgcgcagatctccgtggagatgg
   1081 ---------+---------+---------+---------+---------+---------+ 1140
        ttctcctctagcggcttgacgtccgcgtctaggtcatgcgcgtctagaggcacctctacc

a       K  R  R  S  P  N  C  R  R  R  S  S  T  R  R  S  P  W  R  W   -
b        R  G  D  R  R  T  A  G  A  D  P  V  R  A  D  L  R  G  D  G  -
c         E  E  I  A  E  L  Q  A  Q  I  Q  Y  A  Q  I  S  V  E  M  D -

        acgtgaccaagcccgacctttccgccgcgctcaaggacatccgcgcgcagtacgagaagc
   1141 ---------+---------+---------+---------+---------+---------+ 1200
        tgcactggttcgggctggaaaggcggcgcgagttcctgtaggcgcgcgtcatgctcttcg

a       T  *  P  S  P  T  F  P  P  R  S  R  T  S  A  R  S  T  R  S   -
b        R  D  Q  A  R  P  F  R  R  A  Q  G  H  P  R  A  V  R  E  A  -
c         V  T  K  P  D  L  S  A  A  L  K  D  I  R  A  Q  Y  E  K  L -

        tggccgccaagaacatgcagaacgctgaggaatggttcaagagccgcttcacggtgctga
   1201 ---------+---------+---------+---------+---------+---------+ 1260
        accggcggttcttgtacgtcttgcgactccttaccaagttctcggcgaagtgccacgact
```

Figure 7-5

```
a        W  P  P  R  T  C  R  T  L  R  N  G  S  R  A  A  S  R  C  *   -
b         G  R  Q  E  H  A  E  R  *  G  M  V  Q  E  P  L  H  G  A  D  -
c          A  A  K  N  M  Q  N  A  E  E  W  F  K  S  R  F  T  V  L  T -

         ccgagagcgccgccaagaacaccgacgccgtgcgcgccgccaaggacgaggtgtcggaga
    1261 ---------+---------+---------+---------+---------+---------+ 1320
         ggctctcgcggcggttcttgtggctgcggcacgcgcggcggttcctgctccacagcctct

a        P  R  A  P  P  R  T  P  T  P  C  A  P  P  R  T  R  C  R  R   -
b         R  E  R  R  Q  E  H  R  R  R  A  R  R  Q  G  R  G  V  G  E  -
c          E  S  A  A  K  N  T  D  A  V  R  A  A  K  D  E  V  S  E  S -

         gccgtcgtctgctcaaggccaagaccctggaaatcgaagcatgccggggcatgaatgaag
    1321 ---------+---------+---------+---------+---------+---------+ 1380
         cggcagcagacgagttccggttctgggacctttagcttcgtacggccccgtacttacttc

a        A  V  V  C  S  R  P  R  P  W  K  S  K  H  A  G  A  *  M  K   -
b         P  S  S  A  Q  G  Q  D  P  G  N  R  S  M  P  G  H  E  *  S  -
c          R  R  L  L  K  A  K  T  L  E  I  E  A  C  R  G  M  N  E  A -

         cgctggagaagcagctgcaggagctggaggacaagcagaacgccgacatcagcgctatgc
    1381 ---------+---------+---------+---------+---------+---------+ 1440
         gcgacctcttcgtcgacgtcctcgacctcctgttcgtcttgcggctgtagtcgcgatacg

a        R  W  R  S  S  C  R  S  W  R  T  S  R  T  P  T  S  A  L  C   -
b         A  G  E  A  A  A  G  A  G  G  Q  A  E  R  R  H  Q  R  Y  A  -
c          L  E  K  Q  L  Q  E  L  E  D  K  Q  N  A  D  I  S  A  M  Q -

         aggtgcggcacggccagaaacacaggggggcggggaactcgagcaagggggggagttggt
    1441 ---------+---------+---------+---------+---------+---------+ 1500
         tccacgccgtgccggtctttgtgtccccccgcccccttgagctcgttcccccccctcaacca

a        R  C  G  T  A  R  N  T  G  G  R  G  T  R  A  R  G  G  V  G   -
b         G  A  A  R  P  E  T  Q  G  G  G  E  L  E  Q  G  G  E  L  V  -
c          V  R  H  G  Q  K  H  R  G  A  G  N  S  S  K  G  G  S  W  C -

         gcgcccagaaagcgaaaccaggggtggtgcggctgcccagctcttagggatagggcttgg
    1501 ---------+---------+---------+---------+---------+---------+ 1560
         cgcgggtctttcgctttggtccccaccacgccgacgggtcgagaatccctatcccgaacc

a        A  P  R  K  R  N  Q  G  W  C  G  C  P  A  L  R  D  R  A  W   -
b         R  P  E  S  E  T  R  G  G  A  A  A  Q  L  L  G  I  G  L  G  -
c          A  Q  K  A  K  P  G  V  V  R  L  P  S  S  *  G  *  G  L  A -
```

Figure 7-6

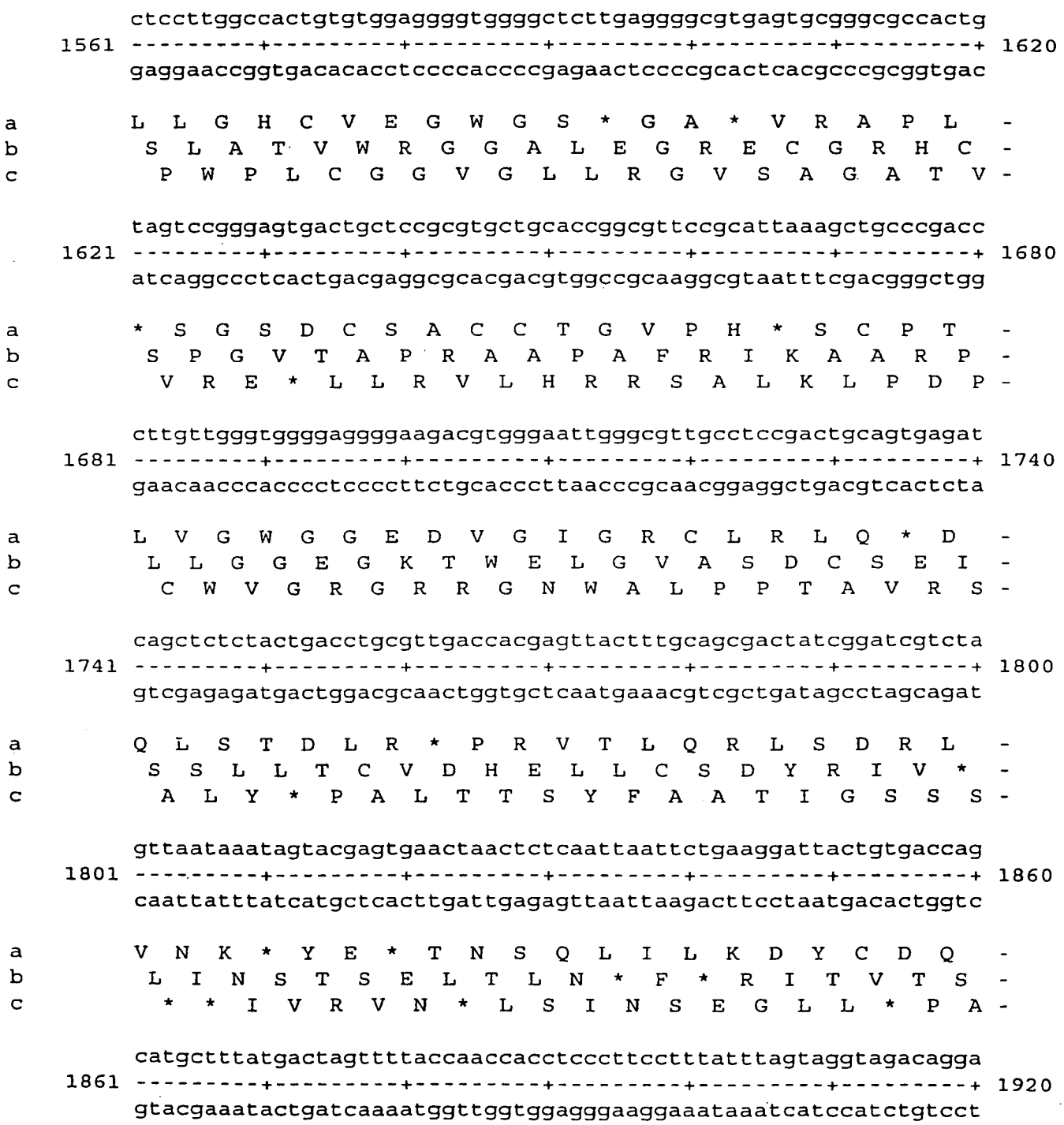
ctccttggccactgtgtggaggggtggggctcttgaggggcgtgagtgcgggcgccactg
1561 ---------+---------+---------+---------+---------+---------+ 1620
gaggaaccggtgacacacctccccaccccgagaactccccgcactcacgcccgcggtgac
a L L G H C V E G W G S * G A * V R A P L -
b S L A T V W R G G A L E G R E C G R H C -
c P W P L C G G V G L L R G V S A G A T V -
tagtccgggagtgactgctccgcgtgctgcaccggcgttccgcattaaagctgcccgacc
1621 ---------+---------+---------+---------+---------+---------+ 1680
atcaggccctcactgacgaggcgcacgacgtggccgcaaggcgtaatttcgacgggctgg
a * S G S D C S A C C T G V P H * S C P T -
b S P G V T A P R A A P A F R I K A A R P -
c V R E * L L R V L H R R S A L K L P D P -
cttgttgggtggggaggggaagacgtgggaattgggcgttgcctccgactgcagtgagat
1681 ---------+---------+---------+---------+---------+---------+ 1740
gaacaacccacccctcccccttctgcaccccttaacccgcaacggaggctgacgtcactcta
a L V G W G G E D V G I G R C L R L Q * D -
b L L G G E G K T W E L G V A S D C S E I -
c C W V G R G R R G N W A L P P T A V R S -
cagctctctactgacctgcgttgaccacgagttactttgcagcgactatcggatcgtcta
1741 ---------+---------+---------+---------+---------+---------+ 1800
gtcgagagatgactggacgcaactggtgctcaatgaaacgtcgctgatagcctagcagat
a Q L S T D L R * P R V T L Q R L S D R L -
b S S L L T C V D H E L L C S D Y R I V * -
c A L Y * P A L T T S Y F A A T I G S S S -
gttaataaatagtacgagtgaactaactctcaattaattctgaaggattactgtgaccag
1801 ---------+---------+---------+---------+---------+---------+ 1860
caattatttatcatgctcacttgattgagagttaattaagacttcctaatgacactggtc
a V N K * Y E * T N S Q L I L K D Y C D Q -
b L I N S T S E L T L N * F * R I T V T S -
c * * I V R V N * L S I N S E G L L * P A -
catgctttatgactagttttaccaaccacctcccttccttttatttagtaggtagacagga
1861 ---------+---------+---------+---------+---------+---------+ 1920
gtacgaaatactgatcaaaatggttggtggagggaaggaaataaatcatccatctgtcct

Figure 7-7

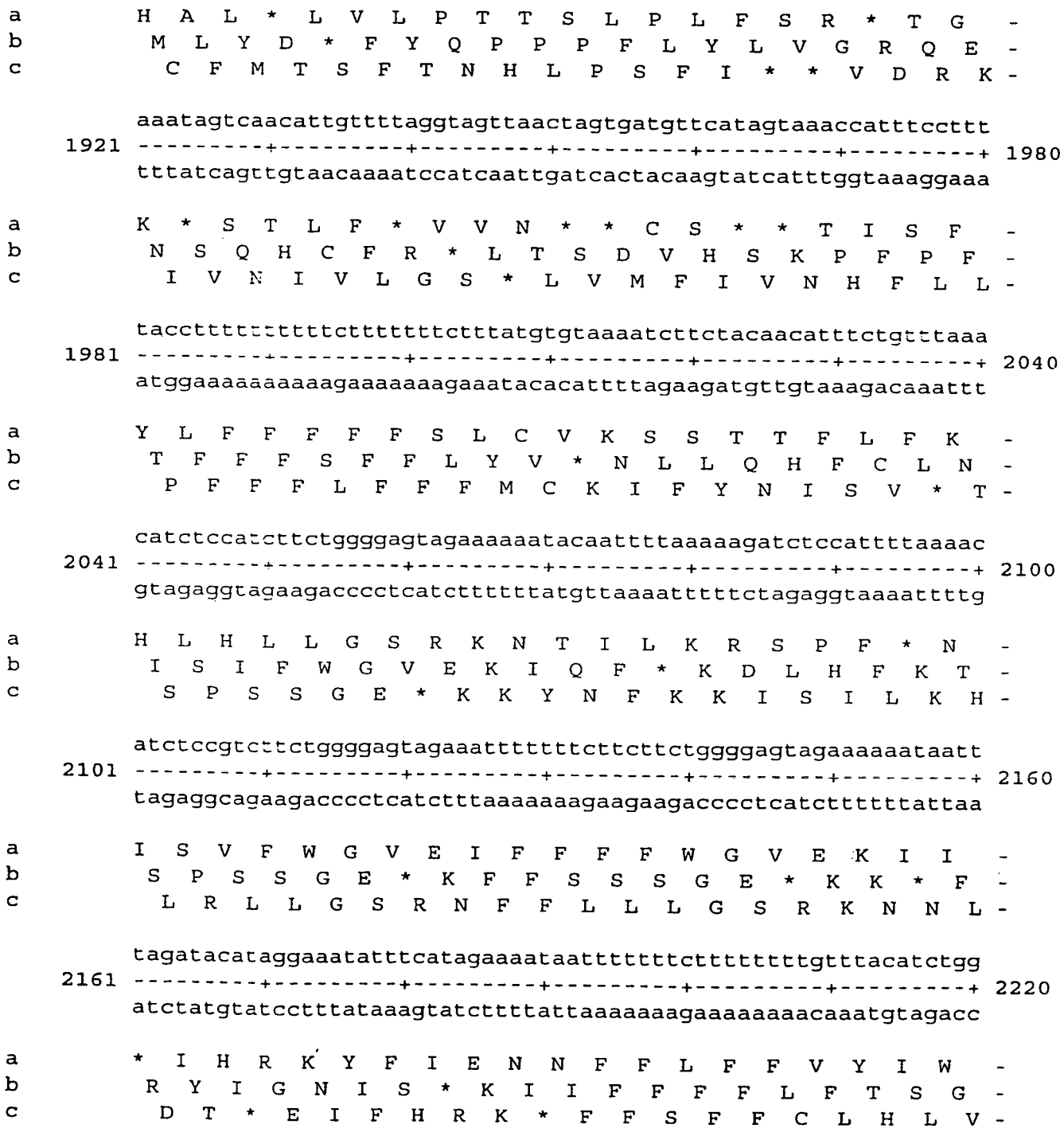
a H A L * L V L P T T S L P L F S R * T G -
b M L Y D * F Y Q P P P F L Y L V G R Q E -
c C F M T S F T N H L P S F I * * V D R K -
aaatagtcaacattgttttaggtagttaactagtgatgttcatagtaaaccatttccttt
1921 ---------+---------+---------+---------+---------+---------+---------+ 1980
tttatcagttgtaacaaaatccatcaattgatcactacaagtatcatttggtaaaggaaa
a K * S T L F * V V N * * C S * * T I S F -
b N S Q H C F R * L T S D V H S K P F P F -
c I V N I V L G S * L V M F I V N H F L L -
tacctttttttttctttttttctttatgtgtaaaatcttctacaacatttctgtttaaa
1981 ---------+---------+---------+---------+---------+---------+---------+ 2040
atggaaaaaaaaaagaaaaaaagaaatacacattttagaagatgttgtaaagacaaattt
a Y L F F F F F S L C V K S S T T F L F K -
b T F F F S F F L Y V * N L L Q H F C L N -
c P F F F L F F F M C K I F Y N I S V * T -
catctccatcttctggggagtagaaaaaatacaattttaaaaagatctccattttaaaac
2041 ---------+---------+---------+---------+---------+---------+---------+ 2100
gtagaggtagaagacccctcatctttttatgttaaaatttttctagaggtaaaattttg
a H L H L L G S R K N T I L K R S P F * N -
b I S I F W G V E K I Q F * K D L H F K T -
c S P S S G E * K K Y N F K K I S I L K H -
atctccgtcttctggggagtagaaatttttttcttcttctggggagtagaaaaaataatt
2101 ---------+---------+---------+---------+---------+---------+---------+ 2160
tagaggcagaagacccctcatctttaaaaaaagaagaagacccctcatctttttattaa
a I S V F W G V E I F F F F W G V E K I I -
b S P S S G E * K F F S S S G E * K K * F -
c L R L L G S R N F F L L L G S R K N N L -
tagatacataggaaatatttcatagaaaataatttttttctttttttttgtttacatctgg
2161 ---------+---------+---------+---------+---------+---------+---------+ 2220
atctatgtatcctttataaagtatcttttattaaaaaaagaaaaaaaacaaatgtagacc
a * I H R K Y F I E N N F F L F F V Y I W -
b R Y I G N I S * K I I F F F F L F T S G -
c D T * E I F H R K * F F S F F C L H L V -

Figure 7-8

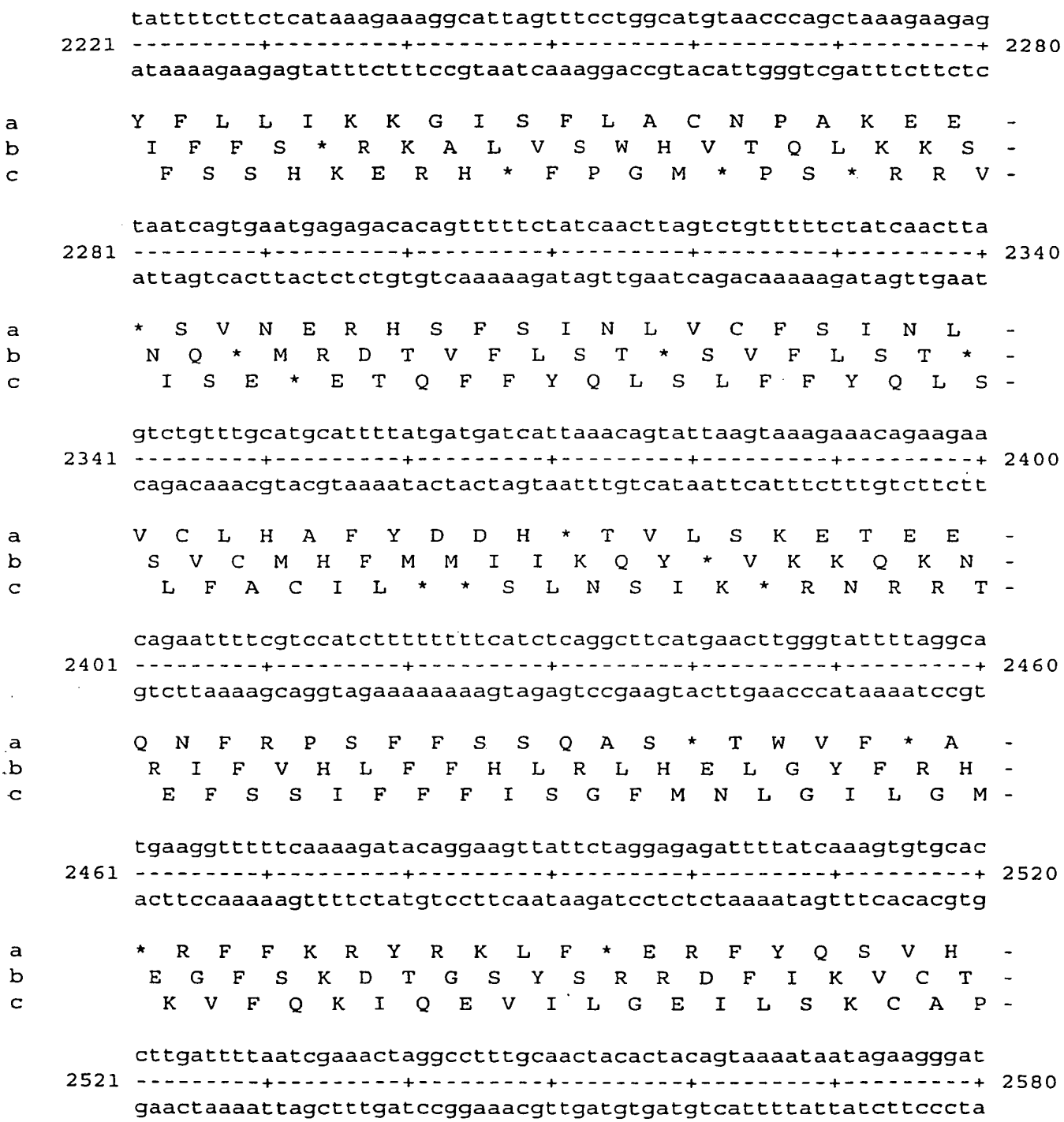
tattttcttctcataaagaaaggcattagtttcctggcatgtaacccagctaaagaagag
2221 ---------+---------+---------+---------+---------+---------+ 2280
ataaaagaagagtatttctttccgtaatcaaaggaccgtacattgggtcgatttcttctc
a Y F L L I K K G I S F L A C N P A K E E -
b I F F S * R K A L V S W H V T Q L K K S -
c F S S H K E R H * F P G M * P S * R R V -
taatcagtgaatgagagacacagtttttctatcaacttagtctgtttttctatcaactta
2281 ---------+---------+---------+---------+---------+---------+ 2340
attagtcacttactctctgtgtcaaaaagatagttgaatcagacaaaaagatagttgaat
a * S V N E R H S F S I N L V C F S I N L -
b N Q * M R D T V F L S T * S V F L S T * -
c I S E * E T Q F F Y Q L S L F F Y Q L S -
gtctgtttgcatgcatttatgatgatcattaaacagtattaagtaaagaaacagaagaa
2341 ---------+---------+---------+---------+---------+---------+ 2400
cagacaaacgtacgtaaaatactactagtaatttgtcataattcatttctttgtcttctt
a V C L H A F Y D D H * T V L S K E T E E -
b S V C M H F M M I I K Q Y * V K K Q K N -
c L F A C I L * * S L N S I K * R N R R T -
cagaattttcgtccatctttttttcatctcaggcttcatgaacttgggtattttaggca
2401 ---------+---------+---------+---------+---------+---------+ 2460
gtcttaaaagcaggtagaaaaaaaagtagagtccgaagtacttgaacccataaaatccgt
a Q N F R P S F F S S Q A S * T W V F * A -
b R I F V H L F F H L R L H E L G Y F R H -
c E F S S I F F F I S G F M N L G I L G M -
tgaaggttttcaaaagatacaggaagttattctaggagagattttatcaaagtgtgcac
2461 ---------+---------+---------+---------+---------+---------+ 2520
acttccaaaaagttttctatgtccttcaataagatcctctctaaaatagtttcacacgtg
a * R F F K R Y R K L F * E R F Y Q S V H -
b E G F S K D T G S Y S R R D F I K V C T -
c K V F Q K I Q E V I L G E I L S K C A P -
cttgattttaatcgaaactaggcctttgcaactacactacagtaaaataatagaagggat
2521 ---------+---------+---------+---------+---------+---------+ 2580
gaactaaaattagctttgatccggaaacgttgatgtgatgtcattttattatcttcccta

Figure 7-9

```
a        L  D  F  N  R  N  *  A  F  A  T  T  L  Q  *  N  N  R  R  D  -
b         L  I  L  I  E  T  R  P  L  Q  L  H  Y  S  K  I  I  E  G  I  -
c          *  F  *  S  K  L  G  L  C  N  Y  T  T  V  K  *  *  K  G  F  -

         ttatgctcggatttttttttgttttatttttgtcttcaaacaggacacgatcaacaaat
    2581 ---------+---------+---------+---------+---------+---------+---------+ 2640
         aatacgagcctaaaaaaaaaacaaaataaaaacagaagtttgtcctgtgctagttgttta

a        L  C  S  D  F  F  F  V  L  F  L  S  S  N  R  T  R  S  T  N  -
b         Y  A  R  I  F  F  L  F  Y  F  C  L  Q  T  G  H  D  Q  Q  I  -
c          M  L  G  F  F  F  C  F  I  F  V  F  K  Q  D  T  I  N  K  L  -

         tagaaaatgaattgaggaccacaaagagtgaaatggcacgatacctaaaagaataccaag
    2641 ---------+---------+---------+---------+---------+---------+---------+ 2700
         atcttttacttaactcctggtgtttctcactttaccgtgctatggattttcttatggttc

a        *  K  M  N  *  G  P  Q  R  V  K  W  H  D  T  *  K  N  T  K  -
b         R  K  *  I  E  D  H  K  E  *  N  G  T  I  P  K  R  I  P  R  -
c          E  N  E  L  R  T  T  K  S  E  M  A  R  Y  L  K  E  Y  Q  D  -

         acctcctcaacgtgaagatggctttggatattgagattgctgcttacaggtgaaaataga
    2701 ---------+---------+---------+---------+---------+---------+---------+ 2760
         tggaggagttgcacttctaccgaaacctataactctaacgacgaatgtccacttttatct

a        T  S  S  T  *  R  W  L  W  I  L  R  L  L  L  T  G  E  N  R  -
b         P  P  Q  R  E  D  G  F  G  Y  *  D  C  C  L  Q  V  K  I  E  -
c          L  L  N  V  K  M  A  L  D  I  E  I  A  A  Y  R  *  K  *  R  -

         ggggcaaagacagcagccattaaaccttaggaagaaaatcagatcccatttaaagttatg
    2761 ---------+---------+---------+---------+---------+---------+---------+ 2820
         ccccgtttctgtcgtcggtaatttggaatccttcttttagtctagggtaaatttcaatac

a        G  A  K  T  A  A  I  K  P  *  E  E  N  Q  I  P  F  K  V  M  -
b         G  Q  R  Q  Q  P  L  N  L  R  K  K  I  R  S  H  L  K  L  C  -
c          G  K  D  S  S  H  *  T  L  G  R  K  S  D  P  I  *  S  Y  V  -

         ttggatcagaaaccttcaataatagtccttttgaaataatgaagtgttagttttggctt
    2821 ---------+---------+---------+---------+---------+---------+---------+ 2880
         aacctagtctttggaagttattatcaggaaaactttattacttcacaatcaaaaaccgaa

a        L  D  Q  K  P  S  I  I  V  L  L  K  *  *  S  V  S  F  W  L  -
b         W  I  R  N  L  Q  *  *  S  F  *  N  N  E  V  L  V  F  G  F  -
c          G  S  E  T  F  N  N  S  P  F  E  I  M  K  C  *  F  L  A  S  -
```

Figure 7-10

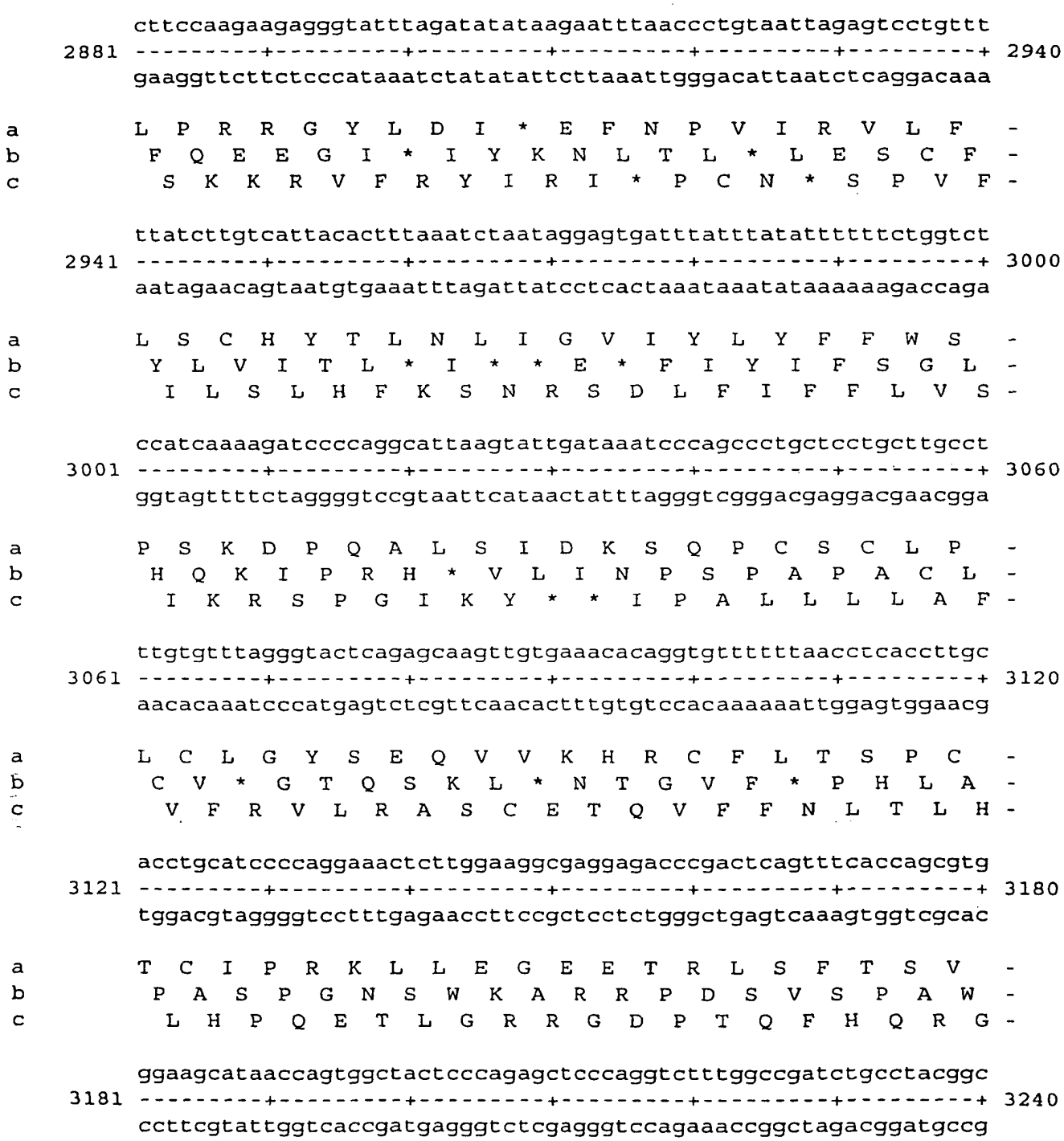
cttccaagaagagggtatttagatatataagaatttaaccctgtaattagagtcctgttt
2881 ---------+---------+---------+---------+---------+---------+ 2940
gaaggttcttctcccataaatctatatattcttaaattgggacattaatctcaggacaaa
a L P R R G Y L D I * E F N P V I R V L F -
b F Q E E G I * I Y K N L T L * L E S C F -
c S K K R V F R Y I R I * P C N * S P V F -
ttatcttgtcattacactttaaatctaataggagtgatttatttatatttttctggtct
2941 ---------+---------+---------+---------+---------+---------+ 3000
aatagaacagtaatgtgaaatttagattatcctcactaaataaatataaaaaagaccaga
a L S C H Y T L N L I G V I Y L Y F F W S -
b Y L V I T L * I * * E * F I Y I F S G L -
c I L S L H F K S N R S D L F I F F L V S -
ccatcaaaagatccccaggcattaagtattgataaatcccagccctgctcctgcttgcct
3001 ---------+---------+---------+---------+---------+---------+ 3060
ggtagttttctaggggtccgtaattcataactatttagggtcgggacgaggacgaacgga
a P S K D P Q A L S I D K S Q P C S C L P -
b H Q K I P R H * V L I N P S P A P A C L -
c I K R S P G I K Y * * I P A L L L L A F -
ttgtgtttagggtactcagagcaagttgtgaaacacaggtgtttttttaacctcaccttgc
3061 ---------+---------+---------+---------+---------+---------+ 3120
aacacaaatcccatgagtctcgttcaacactttgtgtccacaaaaaattggagtggaacg
a L C L G Y S E Q V V K H R C F L T S P C -
b C V * G T Q S K L * N T G V F * P H L A -
c V F R V L R A S C E T Q V F F N L T L H -
acctgcatccccaggaaactcttggaaggcgaggagacccgactcagtttcaccagcgtg
3121 ---------+---------+---------+---------+---------+---------+ 3180
tggacgtaggggtcctttgagaaccttccgctcctctgggctgagtcaaagtggtcgcac
a T C I P R K L L E G E E T R L S F T S V -
b P A S P G N S W K A R R P D S V S P A W -
c L H P Q E T L G R R G D P T Q F H Q R G -
ggaagcataaccagtggctactcccagagctcccaggtctttggccgatctgcctacggc
3181 ---------+---------+---------+---------+---------+---------+ 3240
ccttcgtattggtcaccgatgagggtctcgagggtccagaaaccggctagacggatgccg

Figure 7-11

```
a        G  S  I  T  S  G  Y  S  Q  S  S  Q  V  F  G  R  S  A  Y  G  -
b         E  A  *  P  V  A  T  P  R  A  P  R  S  L  A  D  L  P  T  A  -
c          K  H  N  Q  W  L  L  P  E  L  P  G  L  W  P  I  C  L  R  R -

         ggtttacagaccagctcctatctgatgtccacccgctccttcccgtcctactacaccagc
    3241 ---------+---------+---------+---------+---------+---------+ 3300
         ccaaatgtctggtcgaggatagactacaggtgggcgaggaagggcaggatgatgtggtcg

a        G  L  Q  T  S  S  Y  L  M  S  T  R  S  F  P  S  Y  Y  T  S  -
b         V  Y  R  P  A  P  I  *  C  P  P  A  P  S  R  P  T  T  P  A  -
c          F  T  D  Q  L  L  S  D  V  H  P  L  L  P  V  L  L  H  Q  P -

         catgtccaagaggagcagaccgaagtggaggaaaccattgaggcgtctaaggctgaggaa
    3301 ---------+---------+---------+---------+---------+---------+ 3360
         gtacaggttctcctcgtctggcttcacctcctttggtaactccgcagattccgactcctt

a        H  V  Q  E  E  Q  T  E  V  E  E  T  I  E  A  S  K  A  E  E  -
b         M  S  K  R  S  R  P  K  W  R  K  P  L  R  R  L  R  L  R  K  -
c          C  P  R  G  A  D  R  S  G  G  N  H  *  G  V  *  G  *  G  S -

         gccaaggatgagccccctctgaaggagaagccgaggaggaggagaaggacaaggaagag
    3361 ---------+---------+---------+---------+---------+---------+ 3420
         cggttcctactcgggggagacttcctcttcggctcctcctcctcttcctgttccttctc

a        A  K  D  E  P  P  S  E  G  E  A  E  E  E  E  K  D  K  E  E  -
b         P  R  M  S  P  P  L  K  E  K  P  R  R  R  R  R  T  R  K  R  -
c          Q  G  *  A  P  L  *  R  R  S  R  G  G  G  E  G  Q  G  R  G -

         gccgaggaagaggaggcagctgaagaggaagaaggtatgataagaaaaaaccccctgcaac
    3421 ---------+---------+---------+---------+---------+---------+ 3480
         cggctccttctcctccgtcgacttctccttcttccatactattctttttggggacgttg

a        A  E  E  E  E  A  A  E  E  E  E  G  M  I  R  K  N  P  C  N  -
b         P  R  K  R  R  Q  L  K  R  K  K  V  *  *  E  K  T  P  A  T  -
c          R  G  R  G  G  S  *  R  G  R  R  Y  D  K  K  K  P  L  Q  L -

         ttcaagtgtaaactgggtgtggagatttgttaggaggtggataagacaaatgaagccttg
    3481 ---------+---------+---------+---------+---------+---------+ 3540
         aagttcacatttgacccacacctctaaacaatcctccacctattctgtttacttcggaac

a        F  K  C  K  L  G  V  E  I  C  *  E  V  D  K  T  N  E  A  L  -
b         S  S  V  N  W  V  W  R  F  V  R  R  W  I  R  Q  M  K  P  C  -
c          Q  V  *  T  G  C  G  D  L  L  G  G  G  *  D  K  *  S  L  A -
```

Figure 7-12

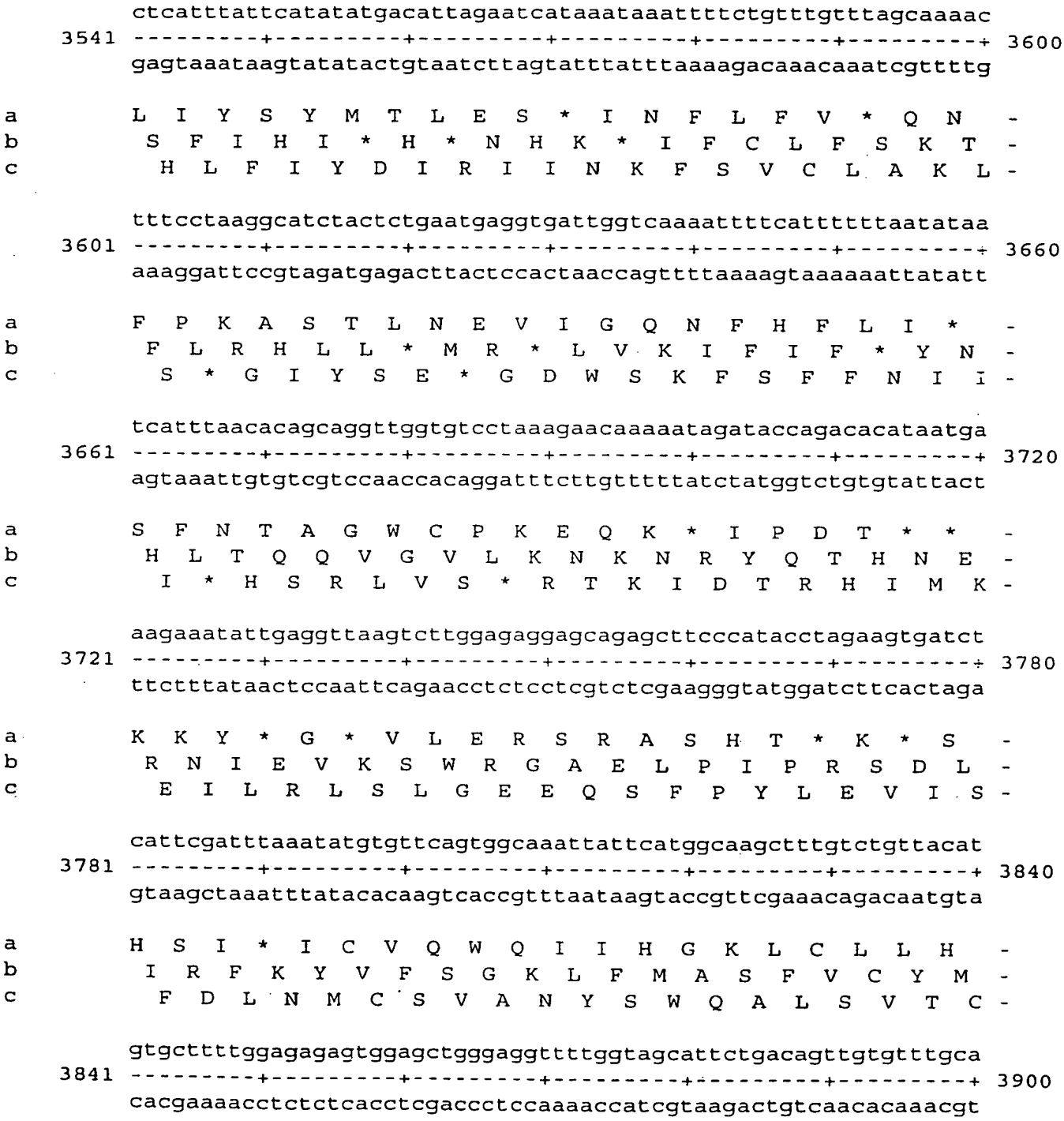
ctcatttattcatatatgacattagaatcataaataaattttctgtttgtttagcaaaac
3541 ---------+---------+---------+---------+---------+---------+ 3600
gagtaaataagtatatactgtaatcttagtatttatttaaaagacaaacaaatcgttttg
a L I Y S Y M T L E S * I N F L F V * Q N -
b S F I H I * H * N H K * I F C L F S K T -
c H L F I Y D I R I I N K F S V C L A K L -
tttcctaaggcatctactctgaatgaggtgattggtcaaaattttcatttttttaatataa
3601 ---------+---------+---------+---------+---------+---------+ 3660
aaaggattccgtagatgagacttactccactaaccagttttaaaagtaaaaaattatatt
a F P K A S T L N E V I G Q N F H F L I * -
b F L R H L L * M R * L V K I F I F * Y N -
c S * G I Y S E * G D W S K F S F F N I I -
tcatttaacacagcaggttggtgtcctaaagaacaaaaatagataccagacacataatga
3661 ---------+---------+---------+---------+---------+---------+ 3720
agtaaattgtgtcgtccaaccacaggatttcttgtttttatctatggtctgtgtattact
a S F N T A G W C P K E Q K * I P D T * * -
b H L T Q Q V G V L K N K N R Y Q T H N E -
c I * H S R L V S * R T K I D T R H I M K -
aagaaatattgaggttaagtcttggagaggagcagagcttcccatacctagaagtgatct
3721 ---------+---------+---------+---------+---------+---------+ 3780
ttctttataactccaattcagaacctctcctcgtctcgaagggtatggatcttcactaga
a K K Y * G * V L E R S R A S H T * K * S -
b R N I E V K S W R G A E L P I P R S D L -
c E I L R L S L G E E Q S F P Y L E V I S -
cattcgatttaaatatgtgttcagtggcaaattattcatggcaagctttgtctgttacat
3781 ---------+---------+---------+---------+---------+---------+ 3840
gtaagctaaatttatacacaagtcaccgtttaataagtaccgttcgaaacagacaatgta
a H S I * I C V Q W Q I I H G K L C L L H -
b I R F K Y V F S G K L F M A S F V C Y M -
c F D L N M C S V A N Y S W Q A L S V T C -
gtgcttttggagagagtggagctgggaggttttggtagcattctgacagttgtgtttgca
3841 ---------+---------+---------+---------+---------+---------+ 3900
cacgaaaacctctctcacctcgaccctccaaaaccatcgtaagactgtcaacacaaacgt

Figure 7-13

```
a        V  L  L  E  R  V  E  L  G  G  F  G  S  I  L  T  V  V  F  A  -
b         C  F  W  R  E  W  S  W  E  V  L  V  A  F  *  Q  L  C  L  Q  -
c          A  F  G  E  S  G  A  G  R  F  W  *  H  S  D  S  C  V  C  K -

         aataaaacctttgcagacatgttttgactggacttaccctggatttgcattttgtacatt
    3901 ---------+---------+---------+---------+---------+---------+ 3960
         ttattttggaaacgtctgtacaaaactgacctgaatgggacctaaacgtaaaacatgtaa

a        N  K  T  F  A  D  M  F  *  L  D  L  P  W  I  C  I  L  Y  I  -
b         I  K  P  L  Q  T  C  F  D  W  T  Y  P  G  F  A  F  C  T  F  -
c          *  N  L  C  R  H  V  L  T  G  L  T  L  D  L  H  F  V  H  F -

         ttctttttatgttaaagctgccaaggaagagtctgaagaagcaaaagaagaagaagaagg
    3961 ---------+---------+---------+---------+---------+---------+ 4020
         aagaaaaatacaatttcgacggttccttctcagacttcttcgttttcttcttcttcttcc

a        F  F  L  C  *  S  C  Q  G  R  V  *  R  S  K  R  R  R  R  R  -
b         S  F  Y  V  K  A  A  K  E  E  S  E  E  A  K  E  E  E  E  G  -
c          L  F  M  L  K  L  P  R  K  S  L  K  K  Q  K  K  K  K  K  E -

         aggtgaaggtgaagaaggagaggaaaccaaagaagctgaagaggaggagaagaaagttga
    4021 ---------+---------+---------+---------+---------+---------+ 4080
         tccacttccacttcttcctctcctttggtttcttcgacttctcctcctcttctttcaact

a        R  *  R  *  R  R  R  G  N  Q  R  S  *  R  G  G  E  E  S  *  -
b         G  E  G  E  E  G  E  E  T  K  E  A  E  E  E  E  K  K  V  E  -
c          V  K  V  K  K  E  R  K  P  K  K  L  K  R  R  R  R  K  L  K -

         aggtgctggggaggaacaagcagctaagaagaaagattgaacccccatttccttaattat
    4081 ---------+---------+---------+---------+---------+---------+ 4140
         tccacgacccctccttgttcgtcgattcttctttctaacttgggggtaaaggaattaata

a        R  C  W  G  G  T  S  S  *  E  E  R  L  N  P  H  F  L  N  Y  -
b         G  A  G  E  E  Q  A  A  K  K  K  D  *  T  P  I  S  L  I  I  -
c          V  L  G  R  N  K  Q  L  R  R  K  I  E  P  P  F  P  *  L  F -

         ttcaggaataattctcccgaaatcaggtcaaccccatcaccaaccaaccaaccagttgag
    4141 ---------+---------+---------+---------+---------+---------+ 4200
         aagtccttattaagagggctttagtccagttggggtagtggttggttggttggtcaactc

a        F  R  N  N  S  P  E  I  R  S  T  P  S  P  T  N  Q  P  V  E  -
b         S  G  I  I  L  P  K  S  G  Q  P  H  H  Q  P  T  N  Q  L  S  -
c          Q  E  *  F  S  R  N  Q  V  N  P  I  T  N  Q  P  T  S  *  V -
```

Figure 7-14

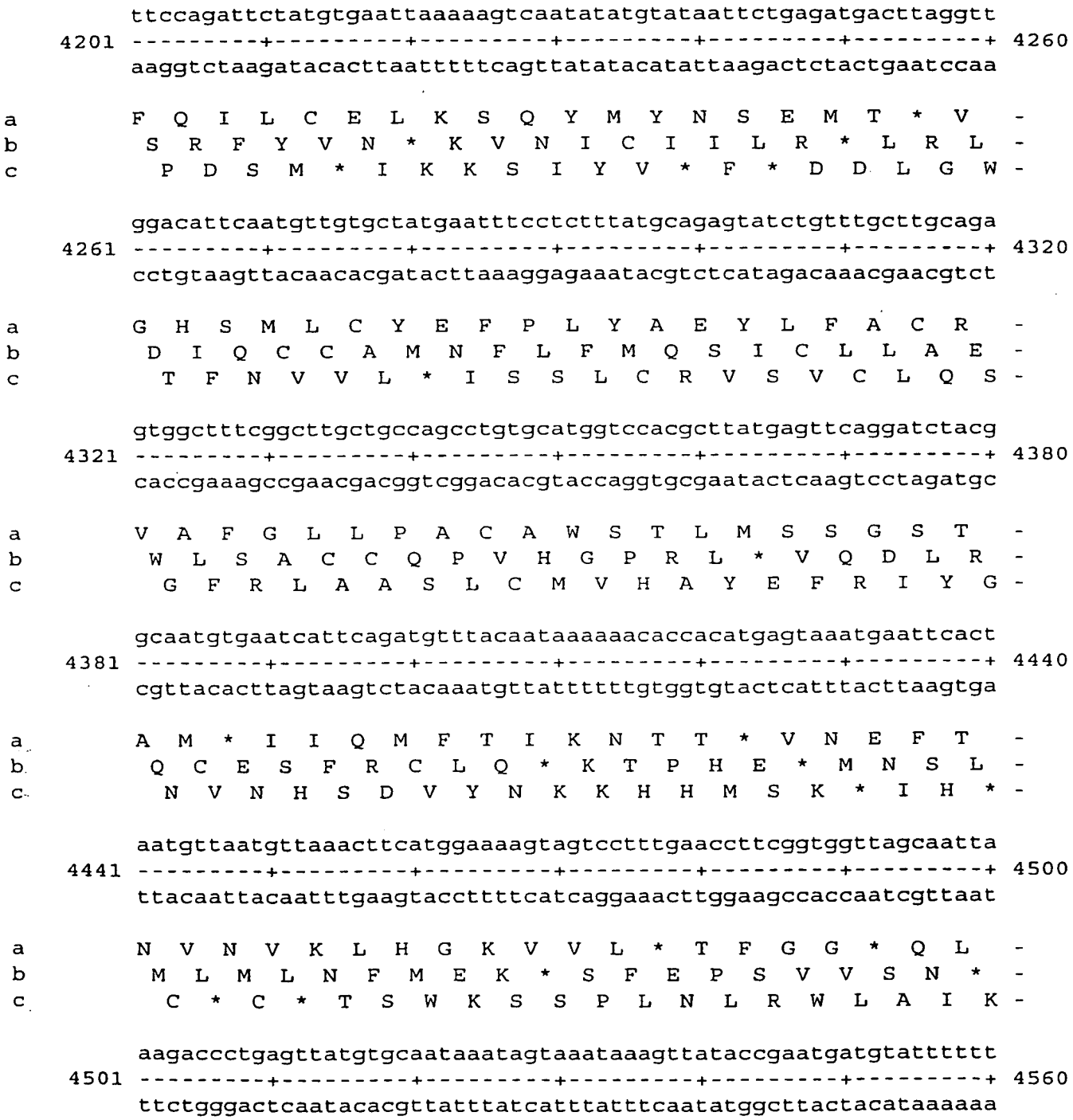
ttccagattctatgtgaattaaaaagtcaatatatgtataattctgagatgacttaggtt
4201 4260
aaggtctaagatacacttaatttttcagttatatacatattaagactctactgaatccaa
a F Q I L C E L K S Q Y M Y N S E M T * V -
b S R F Y V N * K V N I C I I L R * L R L -
c P D S M * I K K S I Y V * F * D D L G W -
ggacattcaatgttgtgctatgaatttcctctttatgcagagtatctgtttgcttgcaga
4261 4320
cctgtaagttacaacacgatacttaaaggagaaatacgtctcatagacaaacgaacgtct
a G H S M L C Y E F P L Y A E Y L F A C R -
b D I Q C C A M N F L F M Q S I C L L A E -
c T F N V V L * I S S L C R V S V C L Q S -
gtggctttcggcttgctgccagcctgtgcatggtccacgcttatgagttcaggatctacg
4321 4380
caccgaaagccgaacgacggtcggacacgtaccaggtgcgaatactcaagtcctagatgc
a V A F G L L P A C A W S T L M S S G S T -
b W L S A C C Q P V H G P R L * V Q D L R -
c G F R L A A S L C M V H A Y E F R I Y G -
gcaatgtgaatcattcagatgtttacaataaaaaacaccacatgagtaaatgaattcact
4381 4440
cgttacacttagtaagtctacaaatgttatttttgtggtgtactcatttacttaagtga
a A M * I I Q M F T I K N T T * V N E F T -
b Q C E S F R C L Q * K T P H E * M N S L -
c N V N H S D V Y N K K H H M S K * I H * -
aatgttaatgttaaacttcatggaaaagtagtcctttgaaccttcggtggttagcaatta
4441 4500
ttacaattacaatttgaagtacctttcatcaggaaacttggaagccaccaatcgttaat
a N V N V K L H G K V V L * T F G G * Q L -
b M L M L N F M E K * S F E P S V V S N * -
c C * C * T S W K S S P L N L R W L A I K -
aagaccctgagttatgtgcaataaatagtaaataaagttataccgaatgatgtatttttt
4501 4560
ttctgggactcaatacacgttatttatcatttatttcaatatggcttactacataaaaaa

Figure 7-15

```
a         K  T  L  S  Y  V  Q  *  I  V  N  K  V  I  P  N  D  V  F  F  -
b          R  P  *  V  M  C  N  K  *  *  I  K  L  Y  R  M  M  Y  F  L  -
c           D  P  E  L  C  A  I  N  S  K  *  S  Y  T  E  *  C  I  F  C -

          gccgtggttgttacctaattaaaataccttaaagatggcaccaatataaagtgtgtgcca
     4561 ---------+---------+---------+---------+---------+---------+ 4620
          cggcaccaacaatggattaattttatggaatttctaccgtggttatatttcacacacggt

a         A  V  V  V  T  *  L  K  Y  L  K  D  G  T  N  I  K  C  V  P  -
b          P  W  L  L  P  N  *  N  T  L  K  M  A  P  I  *  S  V  C  Q  -
c           R  G  C  Y  L  I  K  I  P  *  R  W  H  Q  Y  K  V  C  A  S -

          gtgaactattgacctccaatttttaaaaagccgaaattttaacaattaccaatactttt
     4621 ---------+---------+---------+---------+---------+---------+ 4680
          cacttgataactggaggttaaaaaatttttcggctttaaaattgttaatggttatgaaaa

a         V  N  Y  *  P  P  I  F  *  K  A  E  I  L  T  I  T  N  T  F  -
b          *  T  I  D  L  Q  F  F  K  K  P  K  F  *  Q  L  P  I  L  F  -
c           E  L  L  T  S  N  F  L  K  S  R  N  F  N  N  Y  Q  Y  F  F -

          tt
     4681 -- 4682
          aa

a            -
b            -
c            -
```

Enzymes that do cut:

NONE

Enzymes that do not cut:

NotI

Figure 8-1 Neurofilament M

(Linear) MAP of: hsnfm.gcg check: 3606 from: 1 to: 6236

ID HSNFM standard; DNA; HUM; 6236 BP.
XX
AC Y00067;
XX
NI g35045
XX . . .

With 1 enzymes: NOTI

October 31, 1996 14:29 ..

```
          cagctgctttaagacaaggggtgggggaaggggagggaggcaagaaaagatgagggtggg
        1 ---------+---------+---------+---------+---------+---------+ 60
          gtcgacgaaattctgttccccacccccttcccctccctccgttcttttctactcccaccc

a         Q  L  L  *  D  K  G  W  G  K  G  R  E  A  R  K  D  E  G  G  -
b          S  C  F  K  T  R  G  G  G  R  G  G  R  Q  E  K  M  R  V  G -
c           A  A  L  R  Q  G  V  G  E  G  E  G  G  K  K  R  *  G  W  G -

          ggaggggaaaagagggaatgcaaggggaaggagggaggagacggggagaaggaaagattg
       61 ---------+---------+---------+---------+---------+---------+ 120
          cctccccttttctcccttacgttcccccttcctccctcctctgcccctcttcctttctaac

a         G  G  E  K  R  E  C  K  G  K  E  G  G  D  G  E  K  E  R  L  -
b          E  G  K  R  G  N  A  R  G  R  R  E  E  T  G  R  R  K  D  W -
c           R  G  K  E  G  M  Q  G  E  G  G  R  R  R  G  E  G  K  I  G -

          gaagaaaaggatctccgaggaaggggctgagagaagggcagggtgaactggactaaaggc
      121 ---------+---------+---------+---------+---------+---------+ 180
          cttcttttcctagaggctccttccccgactctcttcccgtcccacttgacctgatttccg

a         E  E  K  D  L  R  G  R  G  *  E  K  G  R  V  N  W  T  K  G  -
b          K  K  R  I  S  E  E  G  A  E  R  R  A  G  *  T  G  L  K  A -
c           R  K  G  S  P  R  K  G  L  R  E  G  Q  G  E  L  D  *  R  P -

          cagagtaggaaggagaagaggggccaaaaaagaaggggatgaaattaagcacagaagatg
      181 ---------+---------+---------+---------+---------+---------+ 240
          gtctcatccttcctcttctccccggtttttttcttcccctactttaattcgtgtcttctac

a         Q  S  R  K  E  K  R  G  Q  K  R  R  G  *  N  *  A  Q  K  M  -
b          R  V  G  R  R  R  G  A  K  K  E  G  D  E  I  K  H  R  R  W -
c           E  *  E  G  E  E  G  P  K  K  K  G  M  K  L  S  T  E  D  G -
```

Figure 8-2

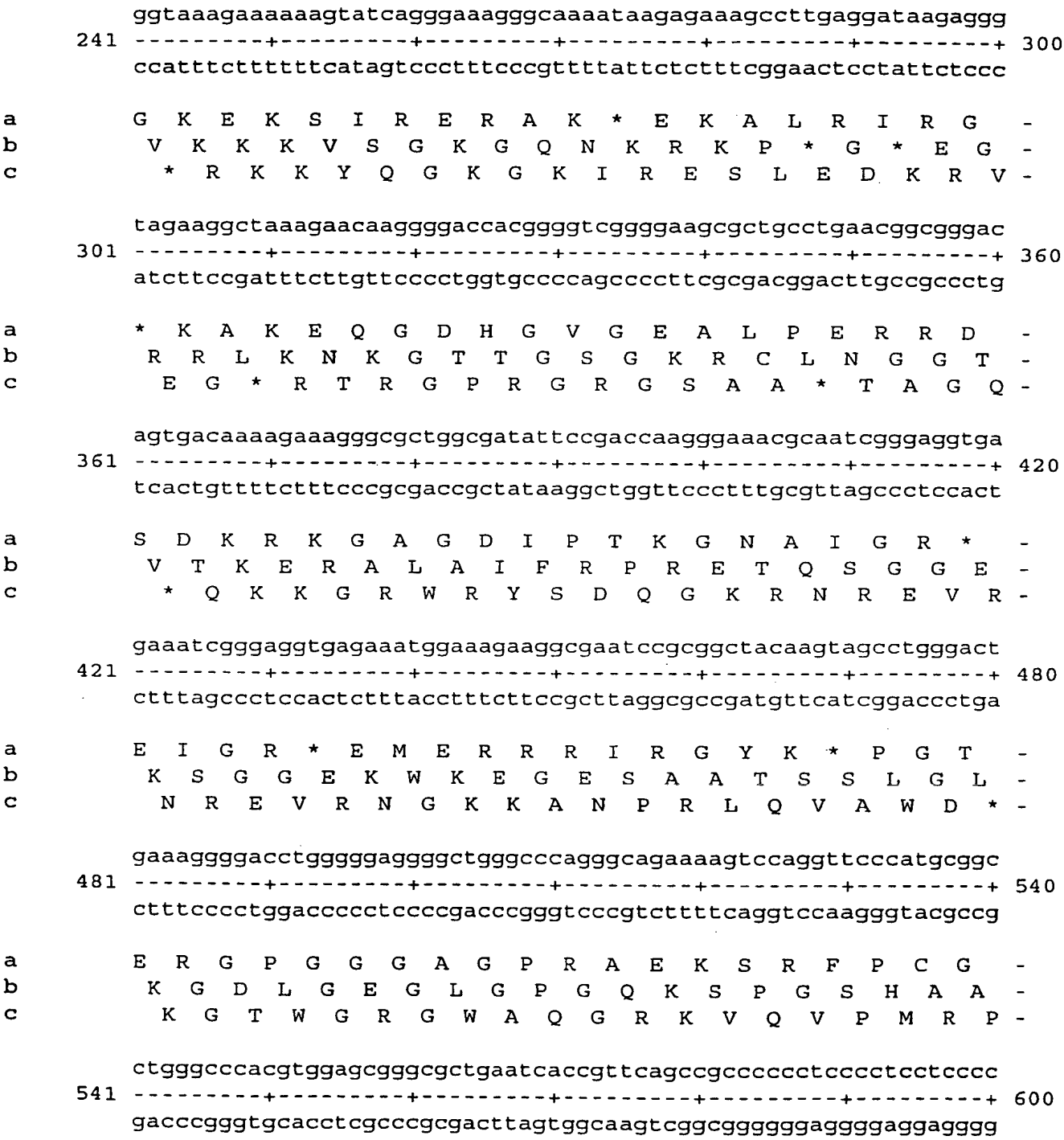
ggtaaagaaaaaagtatcagggaaagggcaaaataagagaaagccttgaggataagaggg
241 ---------+---------+---------+---------+---------+---------+ 300
ccatttctttttcatagtccctttcccgttttattctctttcggaactcctattctccc
a G K E K S I R E R A K * E K A L R I R G -
b V K K K V S G K G Q N K R K P * G * E G -
c * R K K Y Q G K G K I R E S L E D K R V -
tagaaggctaaagaacaaggggaccacggggtcggggaagcgctgcctgaacggcgggac
301 ---------+---------+---------+---------+---------+---------+ 360
atcttccgatttcttgttcccctggtgccccagccccttcgcgacggacttgccgccctg
a * K A K E Q G D H G V G E A L P E R R D -
b R R L K N K G T T G S G K R C L N G G T -
c E G * R T R G P R G R G S A A * T A G Q -
agtgacaaaagaaagggcgctggcgatattccgaccaagggaaacgcaatcgggaggtga
361 ---------+---------+---------+---------+---------+---------+ 420
tcactgttttctttcccgcgaccgctataaggctggttccctttgcgttagccctccact
a S D K R K G A G D I P T K G N A I G R * -
b V T K E R A L A I F R P R E T Q S G G E -
c * Q K K G R W R Y S D Q G K R N R E V R -
gaaatcgggaggtgagaaatggaaagaaggcgaatccgcggctacaagtagcctgggact
421 ---------+---------+---------+---------+---------+---------+ 480
ctttagccctccactctttacctttcttccgcttaggcgccgatgttcatcggaccctga
a E I G R * E M E R R R I R G Y K * P G T -
b K S G G E K W K E G E S A A T S S L G L -
c N R E V R N G K K A N P R L Q V A W D * -
gaaaggggacctgggggaggggctgggcccagggcagaaaagtccaggttcccatgcggc
481 ---------+---------+---------+---------+---------+---------+ 540
ctttcccctggacccccctccccgacccgggtcccgtcttttcaggtccaagggtacgccg
a E R G P G G G A G P R A E K S R F P C G -
b K G D L G E G L G P G Q K S P G S H A A -
c K G T W G R G W A Q G R K V Q V P M R P -
ctgggcccacgtggagcgggcgctgaatcaccgttcagccgcccccctcccctcctcccc
541 ---------+---------+---------+---------+---------+---------+ 600
gacccgggtgcacctcgcccgcgacttagtggcaagtcggcgggggagggggaggagggg

Figure 8-3

```
a        L  G  P  R  G  A  G  A  E  S  P  F  S  R  P  P  P  L  L  P  -
b         W  A  H  V  E  R  A  L  N  H  R  S  A  A  P  L  P  S  S  P  -
c          G  P  T  W  S  G  R  *  I  T  V  Q  P  P  P  S  P  P  P  R -

         gaccggtgcccgcagtccccgcctcctcggccgccgcctccacggggcgggccctggccc
     601 ---------+---------+---------+---------+---------+---------+ 660
         ctggccacgggcgtcaggggcggaggagccggcggcggaggtgccccgcccgggaccggg

a        D  R  C  P  Q  S  P  P  P  R  P  P  P  P  R  G  G  P  W  P  -
b         T  G  A  R  S  P  R  L  L  G  R  R  L  H  G  A  G  P  G  P  .
c          P  V  P  A  V  P  A  S  S  A  A  A  S  T  G  R  A  L  A  R -

         gggaccagcgccgcggctataaatgggctgcggcgaggccggcagaacgctgtgacagcc
     661 ---------+---------+---------+---------+---------+---------+ 720
         ccctggtcgcggcgccgatatttacccgacgccgctccggccgtcttgcgacactgtcgg

a        G  T  S  A  A  A  I  N  G  L  R  R  G  R  Q  N  A  V  T  A  -
b         G  P  A  P  R  L  *  M  G  C  G  E  A  G  R  T  L  *  Q  P  -
c          D  Q  R  R  G  Y  K  W  A  A  A  R  P  A  E  R  C  D  S  H -

         acacgccccaaggcctccaagatgagctacacgttggactcgctgggcaacccgtccgcc
     721 ---------+---------+---------+---------+---------+---------+ 780
         tgtgcggggttccggaggttctactcgatgtgcaacctgagcgacccgttgggcaggcgg

a        T  R  P  K  A  S  K  M  S  Y  T  L  D  S  L  G  N  P  S  A  -
b         H  A  P  R  P  P  R  *  A  T  R  W  T  R  W  A  T  R  P  P  -
c          T  P  Q  G  L  Q  D  E  L  H  V  G  L  A  G  Q  P  V  R  L -

         taccggcgggtaaccgagacccgctcgagcttcagccgcgtcagcggctccccgtccagt
     781 ---------+---------+---------+---------+---------+---------+ 840
         atggccgcccattggctctgggcgagctcgaagtcggcgcagtcgccgaggggcaggtca

a        Y  R  R  V  T  E  T  R  S  S  F  S  R  V  S  G  S  P  S  S  -
b         T  G  G  *  P  R  P  A  R  A  S  A  A  S  A  A  P  R  P  V  -
c          P  A  G  N  R  D  P  L  E  L  Q  P  R  Q  R  L  P  V  Q  W -

         ggcttccgctcgcagtcgtggtcccgcggctcgcccagcaccgtgtcctcctcctataag
     841 ---------+---------+---------+---------+---------+---------+ 900
         ccgaaggcgagcgtcagcaccagggcgccgagcgggtcgtggcacaggaggaggatattc

a        G  F  R  S  Q  S  W  S  R  G  S  P  S  T  V  S  S  S  Y  K  -
b         A  S  A  R  S  R  G  P  A  A  R  P  A  P  C  P  P  P  I  S  -
c          L  P  L  A  V  V  V  P  R  L  A  Q  H  R  V  L  L  L  *  A -
```

Figure 8-4

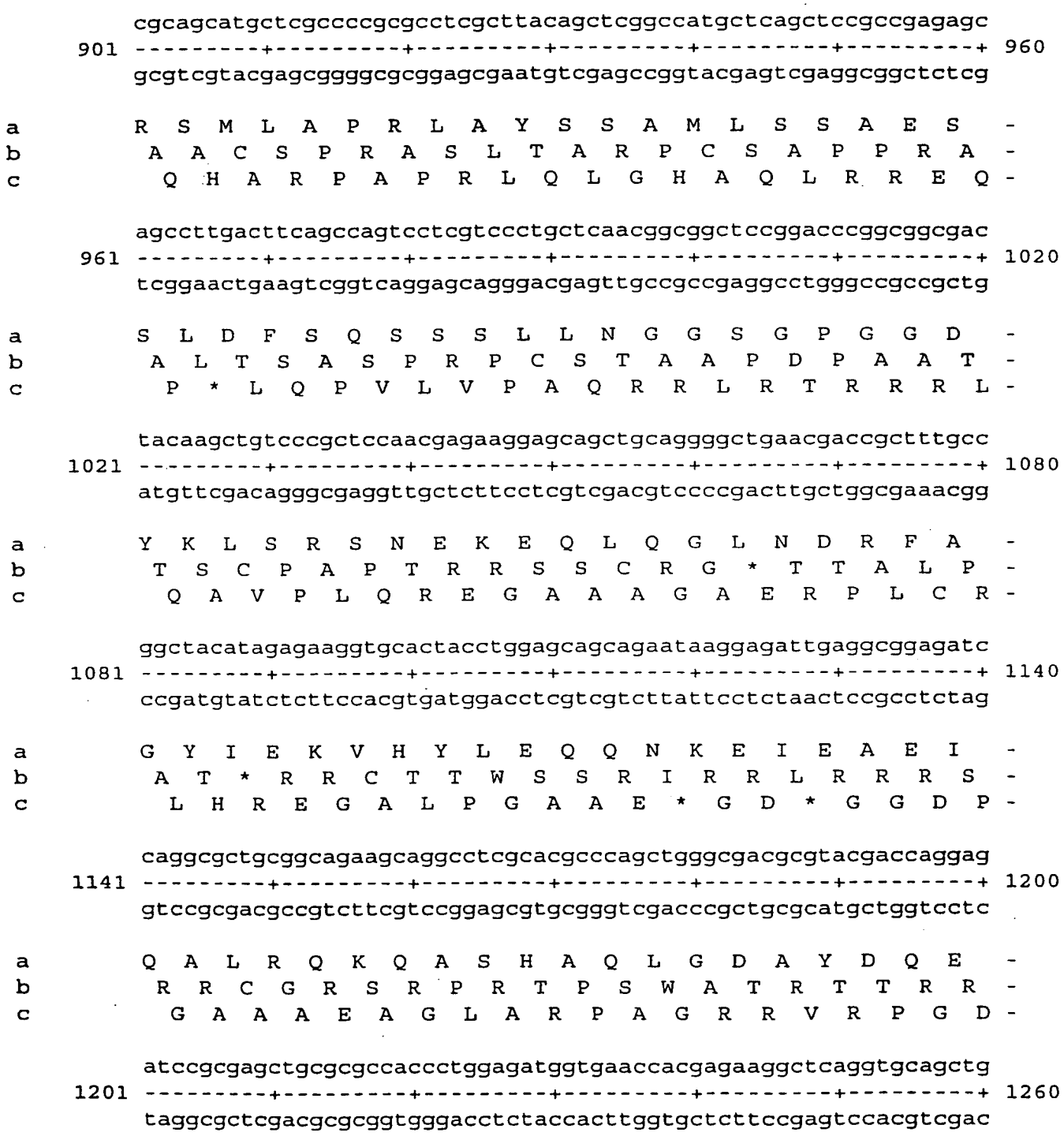
cgcagcatgctcgccccgcgcctcgcttacagctcggccatgctcagctccgccgagagc
901 960
gcgtcgtacgagcggggcgcggagcgaatgtcgagccggtacgagtcgaggcggctctcg
a R S M L A P R L A Y S S A M L S S A E S -
b A A C S P R A S L T A R P C S A P P R A -
c Q H A R P A P R L Q L G H A Q L R R E Q -
agccttgacttcagccagtcctcgtccctgctcaacggcggctccggacccggcggcgac
961 1020
tcggaactgaagtcggtcaggagcagggacgagttgccgccgaggcctgggccgccgctg
a S L D F S Q S S S L L N G G S G P G G D -
b A L T S A S P R P C S T A A P D P A A T -
c P * L Q P V L V P A Q R R L R T R R R L -
tacaagctgtcccgctccaacgagaaggagcagctgcaggggctgaacgaccgctttgcc
1021 1080
atgttcgacagggcgaggttgctcttcctcgtcgacgtccccgacttgctggcgaaacgg
a Y K L S R S N E K E Q L Q G L N D R F A -
b T S C P A P T R R S S C R G * T T A L P -
c Q A V P L Q R E G A A A G A E R P L C R -
ggctacatagagaaggtgcactacctggagcagcagaataaggagattgaggcggagatc
1081 1140
ccgatgtatctcttccacgtgatggacctcgtcgtcttattcctctaactccgcctctag
a G Y I E K V H Y L E Q Q N K E I E A E I -
b A T * R R C T T W S S R I R R L R R R S -
c L H R E G A L P G A A E * G D * G G D P -
caggcgctgcggcagaagcaggcctcgcacgcccagctgggcgacgcgtacgaccaggag
1141 1200
gtccgcgacgccgtcttcgtccggagcgtgcgggtcgacccgctgcgcatgctggtcctc
a Q A L R Q K Q A S H A Q L G D A Y D Q E -
b R R C G R S R P R T P S W A T R T T R R -
c G A A A E A G L A R P A G R R V R P G D -
atccgcgagctgcgcgccaccctggagatggtgaaccacgagaaggctcaggtgcagctg
1201 1260
taggcgctcgacgcgcggtgggacctctaccacttggtgctcttccgagtccacgtcgac

Figure 8-5

```
a        I  R  E  L  R  A  T  L  E  M  V  N  H  E  K  A  Q  V  Q  L  -
b         S  A  S  C  A  P  P  W  R  W  *  T  T  R  R  L  R  C  S  W  -
c          P  R  A  A  R  H  P  G  D  G  E  P  R  E  G  S  G  A  A  G -

         gactcggaccacctggaggaagacatccaccggctcaaggagcgctttgaggaggaggcg
    1261 ---------+---------+---------+---------+---------+---------+---------+ 1320
         ctgagcctggtggacctccttctgtaggtggccgagttcctcgcgaaactcctcctccgc

a        D  S  D  H  L  E  E  D  I  H  R  L  K  E  R  F  E  E  E  A  -
b         T  R  T  T  W  R  K  T  S  T  G  S  R  S  A  L  R  R  R  R  -
c          L  G  P  P  G  G  R  H  P  P  A  Q  G  A  L  *  G  G  G  A -

         cggttgcgggacgacactgaggcggccatccgggcgctgcgcaaagacatcgaggaggcg
    1321 ---------+---------+---------+---------+---------+---------+---------+ 1380
         gccaacgccctgctgtgactccgccggtaggcccgcgacgcgtttctgtagctcctccgc

a        R  L  R  D  D  T  E  A  A  I  R  A  L  R  K  D  I  E  E  A  -
b         G  C  G  T  T  L  R  R  P  S  G  R  C  A  K  T  S  R  R  R  -
c          V  A  G  R  H  *  G  G  H  P  G  A  A  Q  R  H  R  G  G  V -

         tcgctggtcaaggtggagctggacaagaaggtgcagtcgctgcaggatgaggtggccttc
    1381 ---------+---------+---------+---------+---------+---------+---------+ 1440
         agcgaccagttccacctcgacctgttcttccacgtcagcgacgtcctactccaccggaag

a        S  L  V  K  V  E  L  D  K  K  V  Q  S  L  Q  D  E  V  A  F  -
b         R  W  S  R  W  S  W  T  R  R  C  S  R  C  R  M  R  W  P  S  -
c          A  G  Q  G  G  A  G  Q  E  G  A  V  A  A  G  *  G  G  L  P -

         ctgcggagcaaccacgaggaggaggtggccgaccttctggcccagatccaggcatcgcac
    1441 ---------+---------+---------+---------+---------+---------+---------+ 1500
         gacgcctcgttggtgctcctcctccaccggctggaagaccgggtctaggtccgtagcgtg

a        L  R  S  N  H  E  E  E  V  A  D  L  L  A  Q  I  Q  A  S  H  -
b         C  G  A  T  T  R  R  R  W  P  T  F  W  P  R  S  R  H  R  T  -
c          A  E  Q  P  R  G  G  G  G  R  P  S  G  P  D  P  G  I  A  H -

         atcacggtggagcgcaaagactacctgaagacagacatctcgacggcgctgaaggaaatc
    1501 ---------+---------+---------+---------+---------+---------+---------+ 1560
         tagtgccacctcgcgtttctgatggacttctgtctgtagagctgccgcgacttcctttag

a        I  T  V  E  R  K  D  Y  L  K  T  D  I  S  T  A  L  K  E  I  -
b         S  R  W  S  A  K  T  T  *  R  Q  T  S  R  R  R  *  R  K  S  -
c          H  G  G  A  Q  R  L  P  E  D  R  H  L  D  G  A  E  G  N  P -
```

Figure 8-6

```
                cgctcccagctcgaaagccactcagaccagaatatgcaccaggccgaagagtggttcaaa
     1561 ---------+---------+---------+---------+---------+---------+ 1620
                gcgagggtcgagctttcggtgagtctggtcttatacgtggtccggcttctcaccaagttt

a                R  S  Q  L  E  S  H  S  D  Q  N  M  H  Q  A  E  E  W  F  K  -
b                 A  P  S  S  K  A  T  Q  T  R  I  C  T  R  P  K  S  G  S  N  -
c                  L  P  A  R  K  P  L  R  P  E  Y  A  P  G  R  R  V  V  Q  M  -

                tgccgctacgccaagctcaccgaggcggccgagcagaacaaggaggccatccgctccgcc
     1621 ---------+---------+---------+---------+---------+---------+ 1680
                acggcgatgcggttcgagtggctccgccggctcgtcttgttcctccggtaggcgaggcgg

a                C  R  Y  A  K  L  T  E  A  A  E  Q  N  K  E  A  I  R  S  A  -
b                 A  A  T  P  S  S  P  R  R  P  S  R  T  R  R  P  S  A  P  P  -
c                  P  L  R  Q  A  H  R  G  G  R  A  E  Q  G  G  H  P  L  R  Q  -

                aaggaagagatcgccgagtaccggcgccagctgcagtccaagagcatcgagctagagtcg
     1681 ---------+---------+---------+---------+---------+---------+ 1740
                ttccttctctagcggctcatggccgcggtcgacgtcaggttctcgtagctcgatctcagc

a                K  E  E  I  A  E  Y  R  R  Q  L  Q  S  K  S  I  E  L  E  S  -
b                 R  K  R  S  P  S  T  G  A  S  C  S  P  R  A  S  S  *  S  R  -
c                  G  R  D  R  R  V  P  A  P  A  A  V  Q  E  H  R  A  R  V  G  -

                gtgcgcggcaccaaggagtccctggagcggcagctcagcgacatcgaggagcgccacaac
     1741 ---------+---------+---------+---------+---------+---------+ 1800
                cacgcgccgtggttcctcagggacctcgccgtcgagtcgctgtagctcctcgcggtgttg

a                V  R  G  T  K  E  S  L  E  R  Q  L  S  D  I  E  E  R  H  N  -
b                 C  A  A  P  R  S  P  W  S  G  S  S  A  T  S  R  S  A  T  T  -
c                  A  R  H  Q  G  V  P  G  A  A  A  Q  R  H  R  G  A  P  Q  P  -

                cacgacctcagcagctaccaggtaggaaccgcggcctcggccagcctcggccacggccac
     1801 ---------+---------+---------+---------+---------+---------+ 1860
                gtgctggagtcgtcgatggtccatccttggcgccggagccggtcggagccggtgccggtg

a                H  D  L  S  S  Y  Q  V  G  T  A  A  S  A  S  L  G  H  G  H  -
b                 T  T  S  A  A  T  R  *  E  P  R  P  R  P  A  S  A  T  A  T  -
c                  R  P  Q  Q  L  P  G  R  N  R  G  L  G  Q  P  R  P  R  P  R  -

                gccgcgcgcccccgacacttgggctcgtgcccaggcgccctctccgccgcgctccctggt
     1861 ---------+---------+---------+---------+---------+---------+ 1920
                cggcgcgcgggggctgtgaacccgagcacgggtccgcgggagaggcggcgcgagggacca
```

Figure 8-7

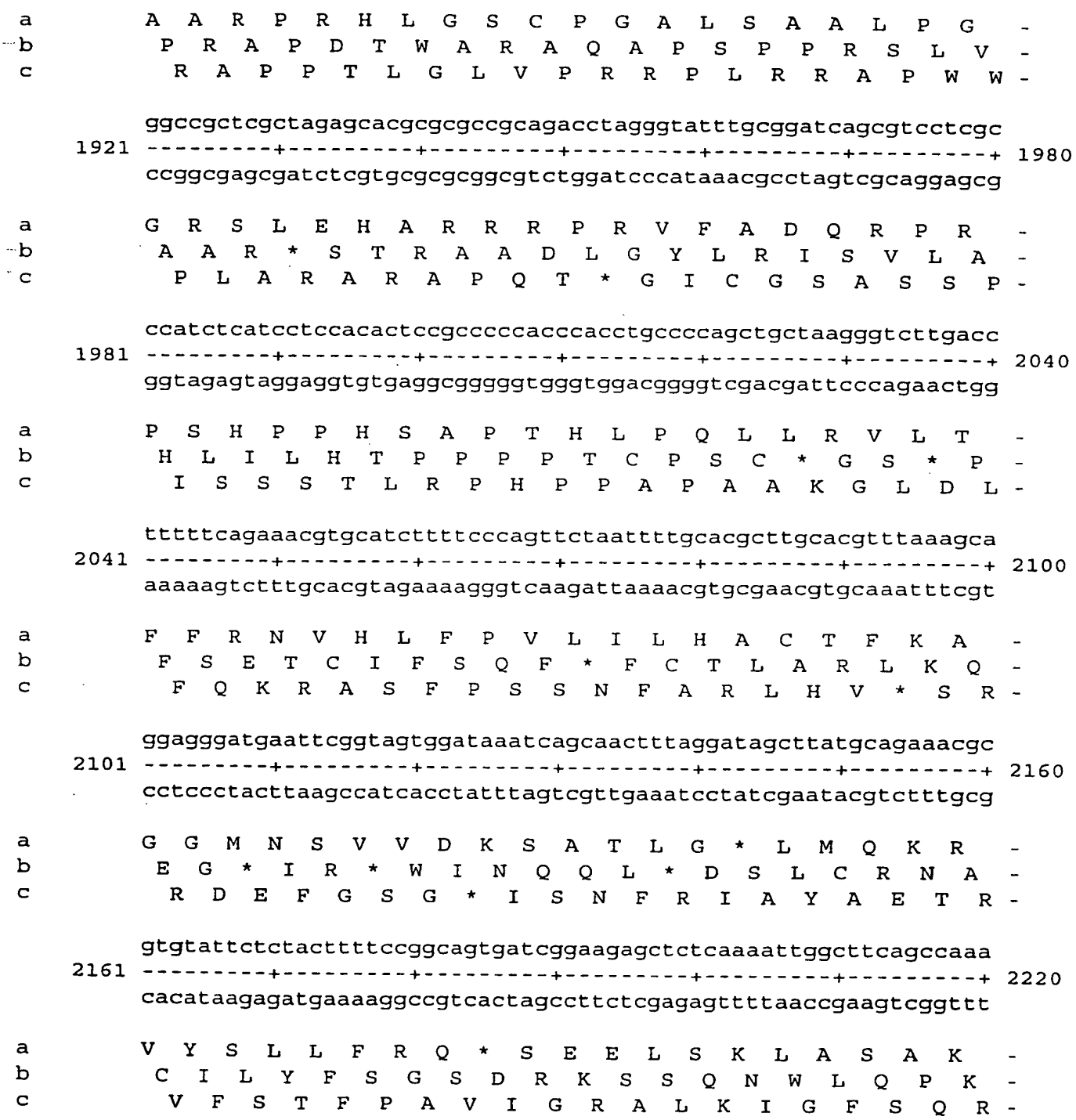
a A A R P R H L G S C P G A L S A A L P G -
b P R A P D T W A R A Q A P S P P R S L V -
c R A P P T L G L V P R R P L R R A P W W -
ggccgctcgctagagcacgcgcgccgcagacctagggtatttgcggatcagcgtcctcgc
1921 ---------+---------+---------+---------+---------+---------+---------+ 1980
ccggcgagcgatctcgtgcgcgcggcgtctggatcccataaacgcctagtcgcaggagcg
a G R S L E H A R R R P R V F A D Q R P R -
b A A R * S T R A A D L G Y L R I S V L A -
c P L A R A R A P Q T * G I C G S A S S P -
ccatctcatcctccacactccgcccccacccacctgccccagctgctaagggtcttgacc
1981 ---------+---------+---------+---------+---------+---------+---------+ 2040
ggtagagtaggaggtgtgaggcgggggtgggtggacggggtcgacgattcccagaactgg
a P S H P P H S A P T H L P Q L L R V L T -
b H L I L H T P P P P T C P S C * G S * P -
c I S S S T L R P H P P A P A A K G L D L -
tttttcagaaacgtgcatcttttcccagttctaattttgcacgcttgcacgtttaaagca
2041 ---------+---------+---------+---------+---------+---------+---------+ 2100
aaaaagtctttgcacgtagaaaagggtcaagattaaaacgtgcgaacgtgcaaatttcgt
a F F R N V H L F P V L I L H A C T F K A -
b F S E T C I F S Q F * F C T L A R L K Q -
c F Q K R A S F P S S N F A R L H V * S R -
ggagggatgaattcggtagtggataaatcagcaactttaggatagcttatgcagaaacgc
2101 ---------+---------+---------+---------+---------+---------+---------+ 2160
cctccctacttaagccatcacctatttagtcgttgaaatcctatcgaatacgtctttgcg
a G G M N S V V D K S A T L G * L M Q K R -
b E G * I R * W I N Q Q L * D S L C R N A -
c R D E F G S G * I S N F R I A Y A E T R -
gtgtattctctacttttccggcagtgatcggaagagctctcaaaattggcttcagccaaa
2161 ---------+---------+---------+---------+---------+---------+---------+ 2220
cacataagagatgaaaaggccgtcactagccttctcgagagttttaaccgaagtcggttt
a V Y S L L F R Q * S E E L S K L A S A K -
b C I L Y F S G S D R K S S Q N W L Q P K -
c V F S T F P A V I G R A L K I G F S Q R -

Figure 8-8

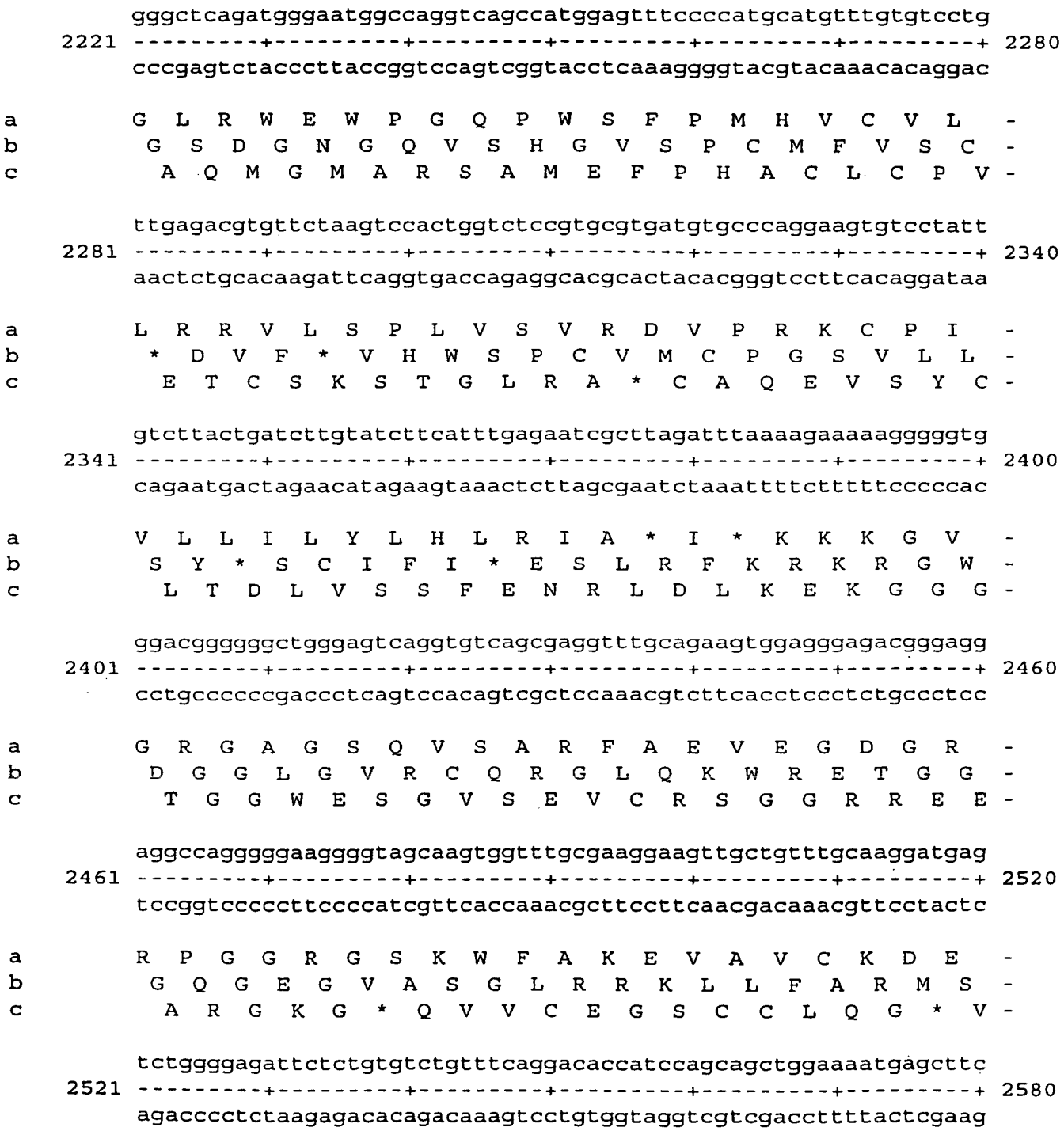
gggctcagatgggaatggccaggtcagccatggagtttccccatgcatgtttgtgtcctg
2221 2280
cccgagtctacccttaccggtccagtcggtacctcaaaggggtacgtacaaacacaggac
a G L R W E W P G Q P W S F P M H V C V L -
b G S D G N G Q V S H G V S P C M F V S C -
c A Q M G M A R S A M E F P H A C L C P V -
ttgagacgtgttctaagtccactggtctccgtgcgtgatgtgcccaggaagtgtcctatt
2281 2340
aactctgcacaagattcaggtgaccagaggcacgcactacacgggtccttcacaggataa
a L R R V L S P L V S V R D V P R K C P I -
b * D V F * V H W S P C V M C P G S V L L -
c E T C S K S T G L R A * C A Q E V S Y C -
gtcttactgatcttgtatcttcatttgagaatcgcttagatttaaaagaaaaaggggtg
2341 2400
cagaatgactagaacatagaagtaaactcttagcgaatctaaattttcttttccccac
a V L L I L Y L H L R I A * I * K K K G V -
b S Y * S C I F I * E S L R F K R K R G W -
c L T D L V S S F E N R L D L K E K G G G -
ggacgggggctgggagtcaggtgtcagcgaggtttgcagaagtggagggagacgggagg
2401 2460
cctgccccccgaccctcagtccacagtcgctccaaacgtcttcacctccctctgccctcc
a G R G A G S Q V S A R F A E V E G D G R -
b D G G L G V R C Q R G L Q K W R E T G G -
c T G G W E S G V S E V C R S G G R R E E -
aggccaggggaaggggtagcaagtggtttgcgaaggaagttgctgtttgcaaggatgag
2461 2520
tccggtccccctttccccatcgttcaccaaacgcttccttcaacgacaaacgttcctactc
a R P G G R G S K W F A K E V A V C K D E -
b G Q G E G V A S G L R R K L L F A R M S -
c A R G K G * Q V V C E G S C C L Q G * V -
tctggggagattctctgtgtctgtttcaggacaccatccagcagctggaaaatgagcttc
2521 2580
agacccctctaagagacacagacaaagtcctgtggtaggtcgtcgacctttactcgaag

Figure 8-9

```
a        S  G  E  I  L  C  V  C  F  R  T  P  S  S  S  W  K  M  S  F  -
b         L  G  R  F  S  V  S  V  S  G  H  H  P  A  A  G  K  *  A  S  -
c          W  G  D  S  L  C  L  F  Q  D  T  I  Q  Q  L  E  N  E  L  R -

         ggggcacaaagtgggaaatggctcgtcatttgcgcgaataccaggacctcctcaacgtca
    2581 ---------+---------+---------+---------+---------+---------+ 2640
         ccccgtgtttcacccttaccgagcagtaaacgcgcttatggtcctggaggagttgcagt

a        G  A  Q  S  G  K  W  L  V  I  C  A  N  T  R  T  S  S  T  S  -
b         G  H  K  V  G  N  G  S  S  F  A  R  I  P  G  P  P  Q  R  Q  -
c          G  T  K  W  E  M  A  R  H  L  R  E  Y  Q  D  L  L  N  V  K -

         agatggctctggatatagaaatcgctgcgtacaggtacgatgcttactacgtgcgtggcc
    2641 ---------+---------+---------+---------+---------+---------+ 2700
         tctaccgagacctatatctttagcgacgcatgtccatgctacgaatgatgcacgcaccgg

a        R  W  L  W  I  *  K  S  L  R  T  G  T  M  L  T  T  C  V  A  -
b         D  G  S  G  Y  R  N  R  C  V  Q  V  R  C  L  L  R  A  W  P  -
c          M  A  L  D  I  E  I  A  A  Y  R  Y  D  A  Y  Y  V  R  G  R -

         ggaacactaaccgcagtgcagaggctgttccggcagagcttccaccacttaagttaaagc
    2701 ---------+---------+---------+---------+---------+---------+ 2760
         ccttgtgattggcgtcacgtctccgacaaggccgtctcgaaggtggtgaattcaatttcg

a        G  T  L  T  A  V  Q  R  L  F  R  Q  S  F  H  H  L  S  *  S  -
b         E  H  *  P  Q  C  R  G  C  S  G  Ŕ  A  S  T  T  *  V  K  A  -
c          N  T  N  R  S  A  E  A  V  P  A  E  L  P  P  L  K  L  K  Q -

         aggcagggtgcaggcatcaactcagcacctggttatcttgcttacttaaaaagaaattat
    2761 ---------+---------+---------+---------+---------+---------+ 2820
         tccgtcccacgtccgtagttgagtcgtggaccaatagaacgaatgaatttttctttaata

a        R  Q  G  A  G  I  N  S  A  P  G  Y  L  A  Y  L  K  R  N  Y  -
b         G  R  V  Q  A  S  T  Q  H  L  V  I  L  L  T  *  K  E  I  I  -
c          A  G  C  R  H  Q  L  S  T  W  L  S  C  L  L  K  K  K  L  F -

         tctaaagaattgcaagtgtagttttatctcttttttatgcagctttaaaagaatgaatact
    2821 ---------+---------+---------+---------+---------+---------+ 2880
         agatttcttaacgttcacatcaaaatagagaaaaatacgtcgaaatttttcttacttatga

a        S  K  E  L  Q  V  *  F  Y  L  F  L  C  S  F  K  R  M  N  T  -
b         L  K  N  C  K  C  S  F  I  S  F  Y  A  A  L  K  E  *  I  L  -
c          *  R  I  A  S  V  V  L  S  L  F  M  Q  L  *  K  N  E  Y  * -
```

Figure 8-10

```
          agtagaaacaaaaggtttttgaattacacaaaggaggtgcagattaatctcaatgcacat
     2881 ---------+---------+---------+---------+---------+---------+ 2940
          tcatctttgttttccaaaaacttaatgtgtttcctccacgtctaattagagttacgtgta

a         S  R  N  K  R  F  L  N  Y  T  K  E  V  Q  I  N  L  N  A  H  -
b          V  E  T  K  G  F  *  I  T  Q  R  R  C  R  L  I  S  M  H  M  -
c           *  K  Q  K  V  F  E  L  H  K  G  G  A  D  *  S  Q  C  T  C -

          gcttaaacttttatggaaaaatgttttcaaatgctggaagcatgaacagagttttggtt
     2941 ---------+---------+---------+---------+---------+---------+ 3000
          cgaatttgaaaaataccttttacaaaagtttacgaccttcgtacttgtctcaaaaccaa

a         A  *  T  F  Y  G  K  M  F  S  N  A  G  S  M  N  R  V  L  V  -
b          L  K  L  F  M  E  K  C  F  Q  M  L  E  A  *  T  E  F  W  F  -
c           L  N  F  L  W  K  N  V  F  K  C  W  K  H  E  Q  S  F  G  F -

          tctaatatttcatctagtggtttcagcttttcaaatgtataatgtcaaggacaaacacca
     3001 ---------+---------+---------+---------+---------+---------+ 3060
          agattataaagtagatcaccaaagtcgaaaagtttacatattacagttcctgtttgtggt

a         S  N  I  S  S  S  G  F  S  F  S  N  V  *  C  Q  G  Q  T  P  -
b          L  I  F  H  L  V  V  S  A  F  Q  M  Y  N  V  K  D  K  H  Q  -
c           *  Y  F  I  *  W  F  Q  L  F  K  C  I  M  S  R  T  N  T  R -

          ggacgttctatttctctgtttctctgttatatagcttactattgccatcatctggctgag
     3061 ---------+---------+---------+---------+---------+---------+ 3120
          cctgcaagataaagagacaaagagacaatatatcgaatgataacggtagtagaccgactc

a         G  R  S  I  S  L  F  L  C  Y  I  A  Y  Y  C  H  H  L  A  E  -
b          D  V  L  F  L  C  F  S  V  I  *  L  T  I  A  I  I  W  L  R  -
c           T  F  Y  F  S  V  S  L  L  Y  S  L  L  L  P  S  S  G  *  E -

          aatagatatagaatgatagaatatagatatagttcttttatatatgtagataatttatat
     3121 ---------+---------+---------+---------+---------+---------+ 3180
          ttatctatatcttactatcttatatctatatcaagaaaatatatacatctattaaatata

a         N  R  Y  R  M  I  E  Y  R  Y  S  S  F  I  Y  V  D  N  L  Y  -
b          I  D  I  E  *  *  N  I  D  I  V  L  L  Y  M  *  I  I  Y  M  -
c           *  I  *  N  D  R  I  *  I  *  F  F  Y  I  C  R  *  F  I  C -

          gtattatattttatgctagactgtagtataaattatatcaatatatcatgtatgatatta
     3181 ---------+---------+---------+---------+---------+---------+ 3240
          cataatataaaatacgatctgacatcatatttaatatagttatatagtacatactataat
```

Figure 8-11

```
a        V  L  Y  F  M  L  D  C  S  I  N  Y  I  N  I  S  C  M  I  L  -
b         Y  Y  I  L  C  *  T  V  V  *  I  I  S  I  Y  H  V  *  Y  *  -
c          I  I  F  Y  A  R  L  *  Y  K  L  Y  Q  Y  I  M  Y  D  I  N -

         atctagatctatagatacacatatgtgcatatgcatataaatctagatatatagacacaa
    3241 ---------+---------+---------+---------+---------+---------+ 3300
         tagatctagatatctatgtgtatacacgtatacgtatatttagatctatatatctgtgtt

a        I  *  I  Y  R  Y  T  Y  V  H  M  H  I  N  L  D  I  *  T  Q  -
b         S  R  S  I  D  T  H  M  C  I  C  I  *  I  *  I  Y  R  H  K  -
c          L  D  L  *  I  H  I  C  A  Y  A  Y  K  S  R  Y  I  D  T  N -

         atatatatgatcgttttatagatagtgagataggttataggtctattaactgaagtgacc
    3301 ---------+---------+---------+---------+---------+---------+ 3360
         tatatatactagcaaaatatctatcactctatccaatatccagataattgacttcactgg

a        I  Y  M  I  V  L  *  I  V  R  *  V  I  G  L  L  T  E  V  T  -
b         Y  I  *  S  F  Y  R  *  *  D  R  L  *  V  Y  *  L  K  *  P  -
c          I  Y  D  R  F  I  D  S  E  I  G  Y  R  S  I  N  *  S  D  L -

         ttgctgttgagtaagcgcaaaggacaaaatcgttgattaaaatttttctgctaccaataa
    3361 ---------+---------+---------+---------+---------+---------+ 3420
         aacgacaactcattcgcgtttcctgttttagcaactaattttaaaaagacgatggttatt

a        L  L  L  S  K  R  K  G  Q  N  R  *  L  K  F  F  C  Y  Q  *  -
b         C  C  *  V  S  A  K  D  K  I  V  D  *  N  F  S  A  T  N  K  -
c          A  V  E  *  A  Q  R  T  K  S  L  I  K  I  F  L  L  P  I  R -

         ggtagttataatataacgagaataaattgcatttacagagctatctctcttttcaggaaa
    3421 ---------+---------+---------+---------+---------+---------+ 3480
         ccatcaatattatattgctcttatttaacgtaaatgtctcgatagagagaaaagtccttt

a        G  S  Y  N  I  T  R  I  N  C  I  Y  R  A  I  S  L  F  R  K  -
b         V  V  I  I  *  R  E  *  I  A  F  T  E  L  S  L  F  S  G  K  -
c          *  L  *  Y  N  E  N  K  L  H  L  Q  S  Y  L  S  F  Q  E  S -

         gctgaataactacattaaatagacactttatgataaaaattatcaacaaatttataactc
    3481 ---------+---------+---------+---------+---------+---------+ 3540
         cgacttattgatgtaatttatctgtgaaatactatttttaatagttgtttaaatattgag

a        A  E  *  L  H  *  I  D  T  L  *  *  K  L  S  T  N  L  *  L  -
b         L  N  N  Y  I  K  *  T  L  Y  D  K  N  Y  Q  Q  I  Y  N  S  -
c          *  I  T  T  L  N  R  H  F  M  I  K  I  I  N  K  F  I  T  R -
```

Figure 8-12

```
          gatacacctgaaaatctaaacgtttaagaaagtgactactctcagaaaggctgtttggct
     3541 ---------+---------+---------+---------+---------+---------+ 3600
          ctatgtggacttttagatttgcaaattctttcactgatgagagtctttccgacaaaccga

a         D  T  P  E  N  L  N  V  *  E  S  D  Y  S  Q  K  G  C  L  A  -
b          I  H  L  K  I  *  T  F  K  K  V  T  T  L  R  K  A  V  W  L  -
c           Y  T  *  K  S  K  R  L  R  K  *  L  L  S  E  R  L  F  G  F -

          ttggagtttgggggcgttttgtttatggcttcttgttttttttgttttttgttttttgttttt
     3601 ---------+---------+---------+---------+---------+---------+ 3660
          aacctcaaacccccgcaaaacaaataccgaagaacaaaaaaacaaaaacaaaaacaaaaa

a         L  E  F  G  G  V  L  F  M  A  S  C  F  F  V  F  V  F  V  F  -
b          W  S  L  G  A  F  C  L  W  L  L  V  F  L  F  L  F  L  F  F  -
c           G  V  W  G  R  F  V  Y  G  F  L  F  F  C  F  C  F  C  F  L -

          tgctatttggccactaacaagtttttcagcatattcatgttgtacctaatggatctctac
     3661 ---------+---------+---------+---------+---------+---------+ 3720
          acgataaaccggtgattgttcaaaaagtcgtataagtacaacatggattacctagagatg

a         C  Y  L  A  T  N  K  F  F  S  I  F  M  L  Y  L  M  D  L  Y  -
b          A  I  W  P  L  T  S  F  S  A  Y  S  C  C  T  *  W  I  S  T  -
c           L  F  G  H  *  Q  V  F  Q  H  I  H  V  V  P  N  G  S  L  L -

          tgcagggccaagacttagtagctgggtgtggttagtggactattgggcaaggttagtcat
     3721 ---------+---------+---------+---------+---------+---------+ 3780
          acgtcccggttctgaatcatcgacccacaccaatcacctgataacccgttccaatcagta

a         C  R  A  K  T  *  *  L  G  V  V  S  G  L  L  G  K  V  S  H  -
b          A  G  P  R  L  S  S  W  V  W  L  V  D  Y  W  A  R  L  V  I  -
c           Q  G  Q  D  L  V  A  G  C  G  *  W  T  I  G  Q  G  *  S  L -

          tgtaggggcaactgtctggcagtccaggagaatctttctctgtcactgagtataatgta
     3781 ---------+---------+---------+---------+---------+---------+ 3840
          acatcccccgttgacagaccgtcaggtcctcttagaaagagacagtgactcatattacat

a         C  R  G  Q  L  S  G  S  P  G  E  S  F  S  V  T  E  Y  N  V  -
b          V  G  G  N  C  L  A  V  Q  E  N  L  S  L  S  L  S  I  M  *  -
c           *  G  A  T  V  W  Q  S  R  R  I  F  L  C  H  *  V  *  C  N -

          atatgccagtaagtgatagcaggtattatagtgaattcatagaatattctacttatgtaa
     3841 ---------+---------+---------+---------+---------+---------+ 3900
          tatacggtcattcactatcgtccataatatcacttaagtatcttataagatgaatacatt
```

Figure 8-13

```
a        I  C  Q  *  V  I  A  G  I  I  V  N  S  *  N  I  L  L  M  *  -
b         Y  A  S  K  *  *  Q  V  L  *  *  I  H  R  I  F  Y  L  C  N -
c          M  P  V  S  D  S  R  Y  Y  S  E  F  I  E  Y  S  T  Y  V  I -

         ttctatttattcaaaggtagctaccacaatacccagaatgtaatgaagctcagaaggcct
    3901 ---------+---------+---------+---------+---------+---------+ 3960
         aagataaataagtttccatcgatggtgttatgggtcttacattacttcgagtcttccgga

a        F  Y  L  F  K  G  S  Y  H  N  T  Q  N  V  M  K  L  R  R  P  -
b         S  I  Y  S  K  V  A  T  T  I  P  R  M  *  *  S  S  E  G  L -
c          L  F  I  Q  R  *  L  P  Q  Y  P  E  C  N  E  A  Q  K  A  * -

         agtgaaatttttactatgtcttatgttcttggattttctccttagaaaactcctggaggg
    3961 ---------+---------+---------+---------+---------+---------+ 4020
         tcactttaaaaatgatacagaatacaagaacctaaaagaggaatcttttgaggacctccc

a        S  E  I  F  T  M  S  Y  V  L  G  F  S  P  *  K  T  P  G  G  -
b         V  K  F  L  L  C  L  M  F  L  D  F  L  L  R  K  L  L  E  G -
c          *  N  F  Y  Y  V  L  C  S  W  I  F  S  L  E  N  S  W  R  V -

         tgaagagactagatttagcacatttgcaggaagcatcactgggccactgtatacacaccg
    4021 ---------+---------+---------+---------+---------+---------+ 4080
         acttctctgatctaaatcgtgtaaacgtccttcgtagtgacccggtgacatatgtgtggc

a        *  R  D  *  I  *  H  I  C  R  K  H  H  W  A  T  V  Y  T  P  -
b         E  E  T  R  F  S  T  F  A  G  S  I  T  G  P  L  Y  T  H  R -
c          K  R  L  D  L  A  H  L  Q  E  A  S  L  G  H  C  I  H  T  D -

         acccccaatcacaatatccagtaagattcagaaaaccaaggtggaagctcccaagcttaa
    4081 ---------+---------+---------+---------+---------+---------+ 4140
         tgggggttagtgttataggtcattctaagtcttttggttccaccttcgagggttcgaatt

a        T  P  N  H  N  I  Q  *  D  S  E  N  Q  G  G  S  S  Q  A  *  -
b         P  P  I  T  I  S  S  K  I  Q  K  T  K  V  E  A  P  K  L  K -
c          P  Q  S  Q  Y  P  V  R  F  R  K  P  R  W  K  L  P  S  L  R -

         ggtccaacacaaatttgtcgaggagatcatagaggaaaccaaagtggaggatgagaagtc
    4141 ---------+---------+---------+---------+---------+---------+ 4200
         ccaggttgtgtttaaacagctcctctagtatctcctttggtttcacctcctactcttcag

a        G  P  T  Q  I  C  R  G  D  H  R  G  N  Q  S  G  G  *  E  V  -
b         V  Q  H  K  F  V  E  E  I  I  E  E  T  K  V  E  D  E  K  S -
c          S  N  T  N  L  S  R  R  S  *  R  K  P  K  W  R  M  R  S  Q -
```

Figure 8-14

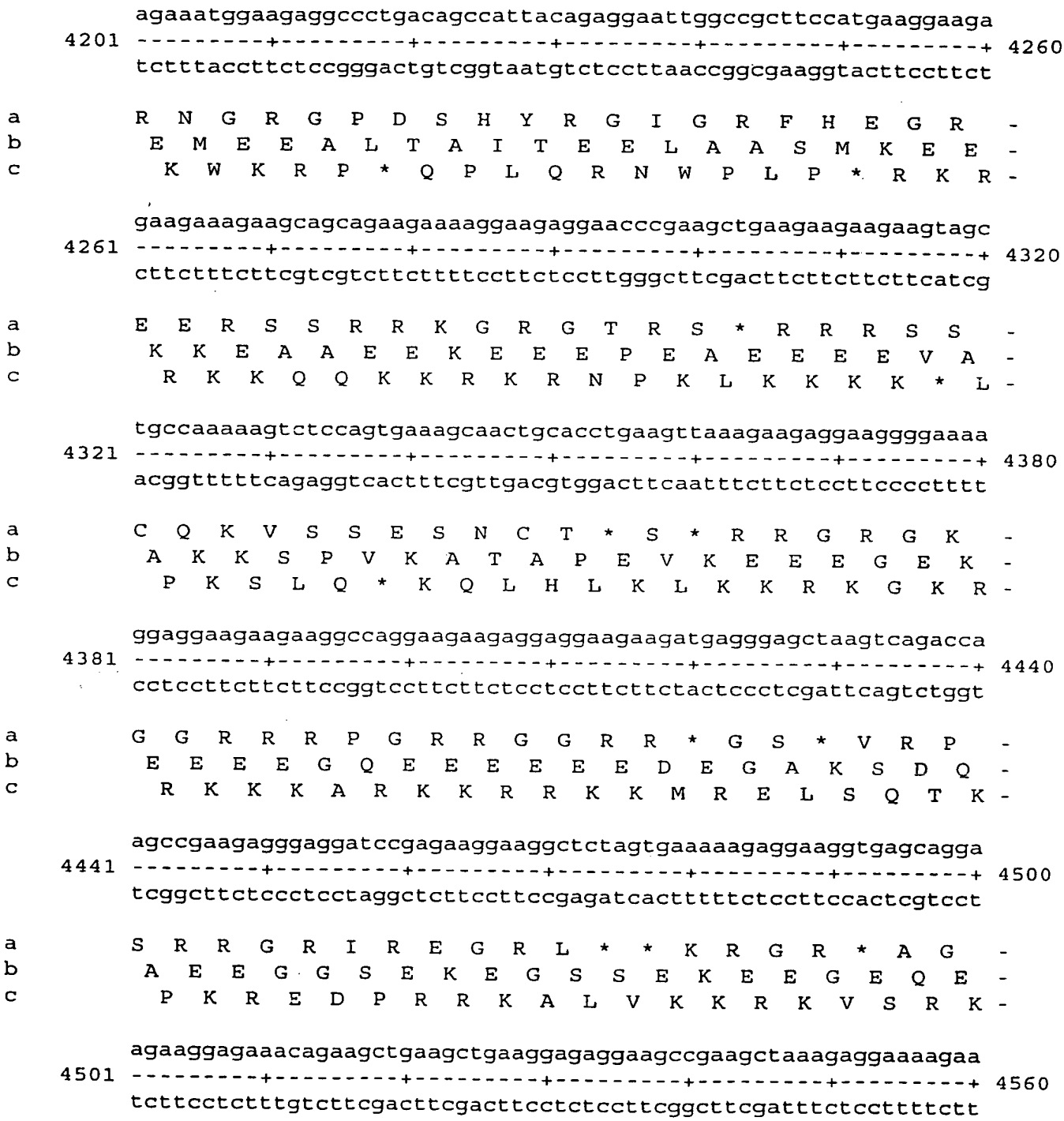
agaaatggaagaggccctgacagccattacagaggaattggccgcttccatgaaggaaga
4201
4260
tctttaccttctccgggactgtcggtaatgtctccttaaccggcgaaggtacttccttct
a R N G R G P D S H Y R G I G R F H E G R -
b E M E E A L T A I T E E L A A S M K E E -
c K W K R P * Q P L Q R N W P L P * R K R -
gaagaaagaagcagcagaagaaaaggaagaggaacccgaagctgaagaagaagaagtagc
4261
4320
cttctttcttcgtcgtcttcttttccttctccttgggcttcgacttcttcttcttcatcg
a E E R S S R R K G R G T R S * R R R S S -
b K K E A A E E K E E E P E A E E E E V A -
c R K K Q Q K K R K R N P K L K K K K * L -
tgccaaaaagtctccagtgaaagcaactgcacctgaagttaaagaagaggaaggggaaaa
4321
4380
acggttttcagaggtcactttcgttgacgtggacttcaatttcttctccttcccctttt
a C Q K V S S E S N C T * S * R R G R G K -
b A K K S P V K A T A P E V K E E E G E K -
c P K S L Q * K Q L H L K L K K R K G K R -
ggaggaagaagaaggccaggaagaagaggaggaagaagatgagggagctaagtcagacca
4381
4440
cctccttcttcttccggtccttcttctcctccttcttctactccctcgattcagtctggt
a G G R R R P G R R G G R R * G S * V R P -
b E E E E G Q E E E E E E D E G A K S D Q -
c R K K K A R K K R R K K M R E L S Q T K -
agccgaagagggaggatccgagaaggaaggctctagtgaaaaagaggaaggtgagcagga
4441
4500
tcggcttctccctcctaggctcttccttccgagatcacttttctccttccactcgtcct
a S R R G R I R E G R L * * K R G R * A G -
b A E E G G S E K E G S S E K E E G E Q E -
c P K R E D P R R K A L V K K R K V S R K -
agaaggagaaacagaagctgaagctgaaggagaggaagccgaagctaaagaggaaaagaa
4501
4560
tcttcctctttgtcttcgacttcgacttcctctccttcggcttcgatttctccttttctt

Figure 8-15

```
a        R  R  R  N  R  S  *  S  *  R  R  G  S  R  S  *  R  G  K  E  -
b         E  G  E  T  E  A  E  A  E  G  E  E  A  E  A  K  E  E  K  K  -
c          K  E  K  Q  K  L  K  L  K  E  R  K  P  K  L  K  R  K  R  K -

         agtggaggaaaagagtgaggaagtggctaccaaggaggagctggtggcagatgccaaggt
    4561 ---------+---------+---------+---------+---------+---------+ 4620
         tcacctcctttctcactccttcaccgatggttcctcctcgaccaccgtctacggttcca

a        S  G  G  K  E  *  G  S  G  Y  Q  G  G  A  G  G  R  C  Q  G  -
b         V  E  E  K  S  E  E  V  A  T  K  E  E  L  V  A  D  A  K  V  -
c          W  R  K  R  V  R  K  W  L  P  R  R  S  W  W  Q  M  P  R  W -

         ggaaaagccagaaaaagccaagtctcctgtgccaaaatcaccagtggaagagaaaggcaa
    4621 ---------+---------+---------+---------+---------+---------+ 4680
         ccttttcggtcttttcggttcagaggacacggttttagtggtcaccttctctttccgtt

a        G  K  A  R  K  S  Q  V  S  C  A  K  I  T  S  G  R  E  R  Q  -
b         E  K  P  E  K  A  K  S  P  V  P  K  S  P  V  E  E  K  G  K  -
c          K  S  Q  K  K  P  S  L  L  C  Q  N  H  Q  W  K  R  K  A  S -

         gtctccctgtgcccaagtcaccagtggaagagaaaggcaagtctcctgtgcccaagtcacc
    4681 ---------+---------+---------+---------+---------+---------+ 4740
         cagaggacacgggttcagtggtcaccttctctttccgttcagaggacacgggttcagtgg

a        V  S  C  A  Q  V  T  S  G  R  E  R  Q  V  S  C  A  Q  V  T  -
b         S  P  V  P  K  S  P  V  E  E  K  G  K  S  P  V  P  K  S  P  -
c          L  L  C  P  S  H  Q  W  K  R  K  A  S  L  L  C  P  S  H  Q -

         agtggaagagaaaggcaagtctcctgtgccgaaatcaccagtggaagagaaaggcaagtc
    4741 ---------+---------+---------+---------+---------+---------+ 4800
         tcaccttctctttccgttcagaggacacggctttagtggtcaccttctctttccgttcag

a        S  G  R  E  R  Q  V  S  C  A  E  I  T  S  G  R  E  R  Q  V  -
b         V  E  E  K  G  K  S  P  V  P  K  S  P  V  E  E  K  G  K  S  -
c          W  K  R  K  A  S  L  L  C  R  N  H  Q  W  K  R  K  A  S  L -

         tcctgtgtcaaaatcaccagtggaagagaaagccaaatctcctgtgccaaaatcaccagt
    4801 ---------+---------+---------+---------+---------+---------+ 4860
         aggacacagttttagtggtcaccttctctttcggtttagaggacacggttttagtggtca

a        S  C  V  K  I  T  S  G  R  E  S  Q  I  S  C  A  K  I  T  S  -
b         P  V  S  K  S  P  V  E  E  K  A  K  S  P  V  P  K  S  P  V  -
c          L  C  Q  N  H  Q  W  K  R  K  P  N  L  L  C  Q  N  H  Q  W -
```

Figure 8-16

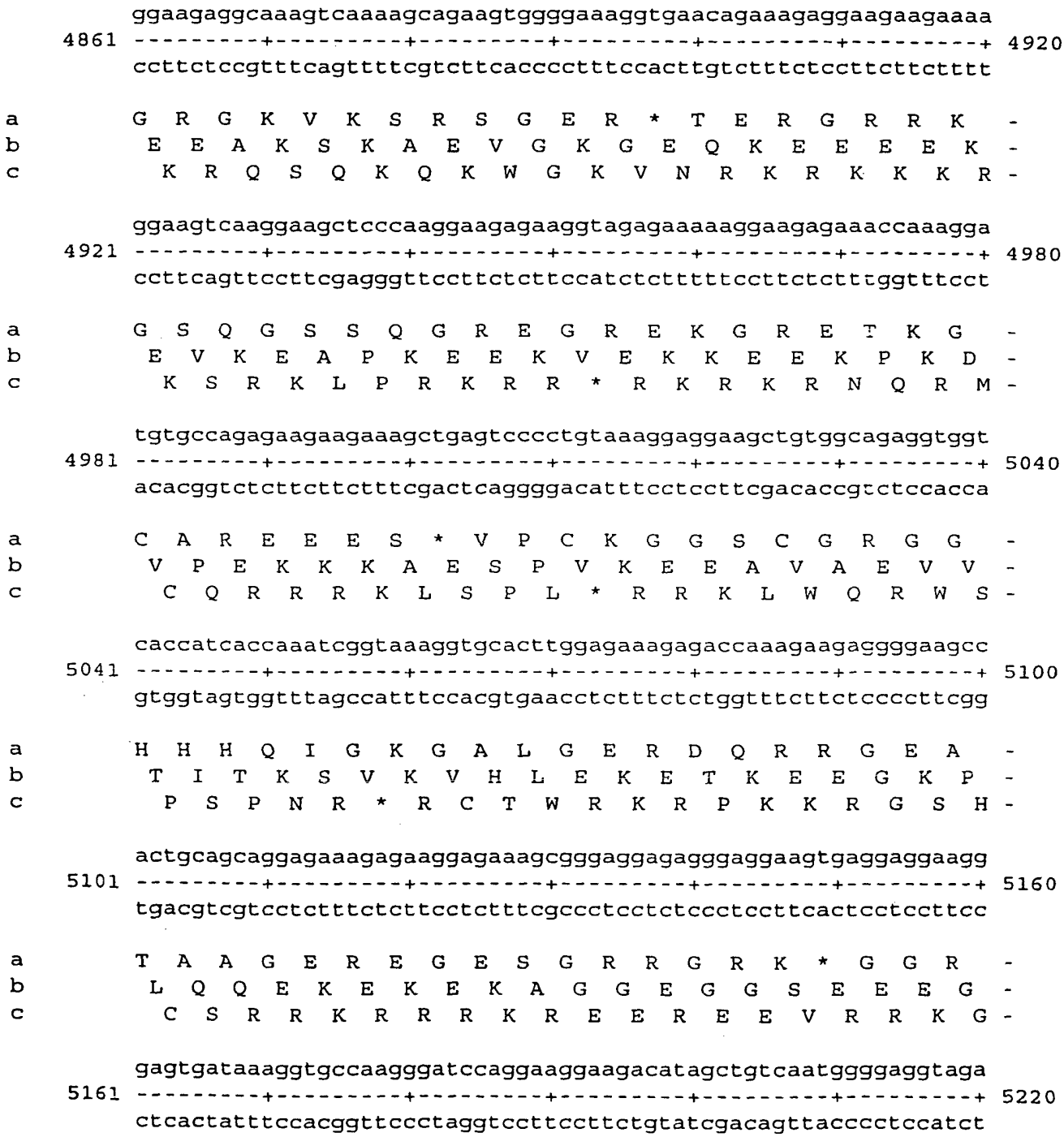
ggaagaggcaaagtcaaaagcagaagtggggaaaggtgaacagaaagaggaagaagaaaa
4861 ---------+---------+---------+---------+---------+---------+ 4920
ccttctccgtttcagttttcgtcttcacccctttccacttgtctttctccttcttctttt
a G R G K V K S R S G E R * T E R G R R K -
b E E A K S K A E V G K G E Q K E E E E K -
c K R Q S Q K Q K W G K V N R K R K K K R -
ggaagtcaaggaagctcccaaggaagagaaggtagagaaaaaggaagagaaaccaaagga
4921 ---------+---------+---------+---------+---------+---------+ 4980
ccttcagttccttcgagggttccttctcttccatctctttttccttctctttggtttcct
a G S Q G S S Q G R E G R E K G R E T K G -
b E V K E A P K E E K V E K K E E K P K D -
c K S R K L P R K R R * R K R K R N Q R M -
tgtgccagagaagaagaaagctgagtcccctgtaaaggaggaagctgtggcagaggtggt
4981 ---------+---------+---------+---------+---------+---------+ 5040
acacggtctcttcttctttcgactcaggggacatttcctccttcgacaccgtctccacca
a C A R E E E S * V P C K G G S C G R G G -
b V P E K K K A E S P V K E E A V A E V V -
c C Q R R R K L S P L * R R K L W Q R W S -
caccatcaccaaatcggtaaaggtgcacttggagaaagagaccaaagaagaggggaagcc
5041 ---------+---------+---------+---------+---------+---------+ 5100
gtggtagtggtttagccatttccacgtgaacctctttctctggtttcttctccccttcgg
a H H H Q I G K G A L G E R D Q R R G E A -
b T I T K S V K V H L E K E T K E E G K P -
c P S P N R * R C T W R K R P K K R G S H -
actgcagcaggagaaagagaaggagaaagcgggaggagagggaggaagtgaggaggaagg
5101 ---------+---------+---------+---------+---------+---------+ 5160
tgacgtcgtcctctttctcttcctctttcgccctcctctccctccttcactcctccttcc
a T A A G E R E G E S G R R G R K * G G R -
b L Q Q E K E K E K A G G E G G S E E E G -
c C S R R K R R R K R E E R E E V R R K G -
gagtgataaaggtgccaagggatccaggaaggaagacatagctgtcaatggggaggtaga
5161 ---------+---------+---------+---------+---------+---------+ 5220
ctcactatttccacggttccctaggtccttccttctgtatcgacagttacccctccatct

Figure 8-17

```
a        E  *  *  R  C  Q  G  I  Q  E  G  R  H  S  C  Q  W  G  G  R  -
b         S  D  K  G  A  K  G  S  R  K  E  D  I  A  V  N  G  E  V  E  -
c          V  I  K  V  P  R  D  P  G  R  K  T  *  L  S  M  G  R  *  K -

         aggaaaagaggaggtagagcaggagaccaaggaaaaaggcagtgggagggaagaggagaa
    5221 ---------+---------+---------+---------+---------+---------+ 5280
         tccttttctcctccatctcgtcctctggttccttttccgtcaccctcccttctcctctt

a        R  K  R  G  G  R  A  G  D  Q  G  K  R  Q  W  E  G  R  G  E  -
b         G  K  E  E  V  E  Q  E  T  K  E  K  G  S  G  R  E  E  E  K  -
c          E  K  R  R  *  S  R  R  P  R  K  K  A  V  G  G  K  R  R  K -

         aggcgttgtcaccaatggcctagacttgagcccagcagatgaaaagaaggggggtgataa
    5281 ---------+---------+---------+---------+---------+---------+ 5340
         tccgcaacagtggttaccggatctgaactcgggtcgtctacttttcttccccccactatt

a        R  R  C  H  Q  W  P  R  L  E  P  S  R  *  K  E  G  G  *  *  -
b         G  V  V  T  N  G  L  D  L  S  P  A  D  E  K  K  G  G  D  K  -
c          A  L  S  P  M  A  *  T  *  A  Q  Q  M  K  R  R  G  V  I  K -

         aagtgaggagaaagtggtggtgaccaaaacggtagaaaaaatcaccagtgaggggggaga
    5341 ---------+---------+---------+---------+---------+---------+ 5400
         ttcactcctctttcaccaccactggttttgccatctttttttagtggtcactcccccctct

a        K  *  G  E  S  G  G  D  Q  N  G  R  K  N  H  Q  *  G  G  R  -
b         S  E  E  K  V  V  V  T  K  T  V  E  K  I  T  S  E  G  G  D  -
c          V  R  R  K  W  W  *  P  K  R  *  K  K  S  P  V  R  G  E  M -

         tggtgctaccaaatacatcactaaatctgtaaccgtcactcaaaaggttgaagagcatga
    5401 ---------+---------+---------+---------+---------+---------+ 5460
         accacgatggtttatgtagtgatttagacattggcagtgagttttccaacttctcgtact

a        W  C  Y  Q  I  H  H  *  I  C  N  R  H  S  K  G  *  R  A  *  -
b         G  A  T  K  Y  I  T  K  S  V  T  V  T  Q  K  V  E  E  H  E  -
c          V  L  P  N  T  S  L  N  L  *  P  S  L  K  R  L  K  S  M  K -

         agagacctttgaggagaaactagtgtctactaaaaaggtagaaaaagtcacttcacacgc
    5461 ---------+---------+---------+---------+---------+---------+ 5520
         tctctggaaactcctctttgatcacagatgatttttccatctttttcagtgaagtgtgcg

a        R  D  L  *  G  E  T  S  V  Y  *  K  G  R  K  S  H  F  T  R  -
b         E  T  F  E  E  K  L  V  S  T  K  K  V  E  K  V  T  S  H  A  -
c          R  P  L  R  R  N  *  C  L  L  K  R  *  K  K  S  L  H  T  P -
```

Figure 8-18

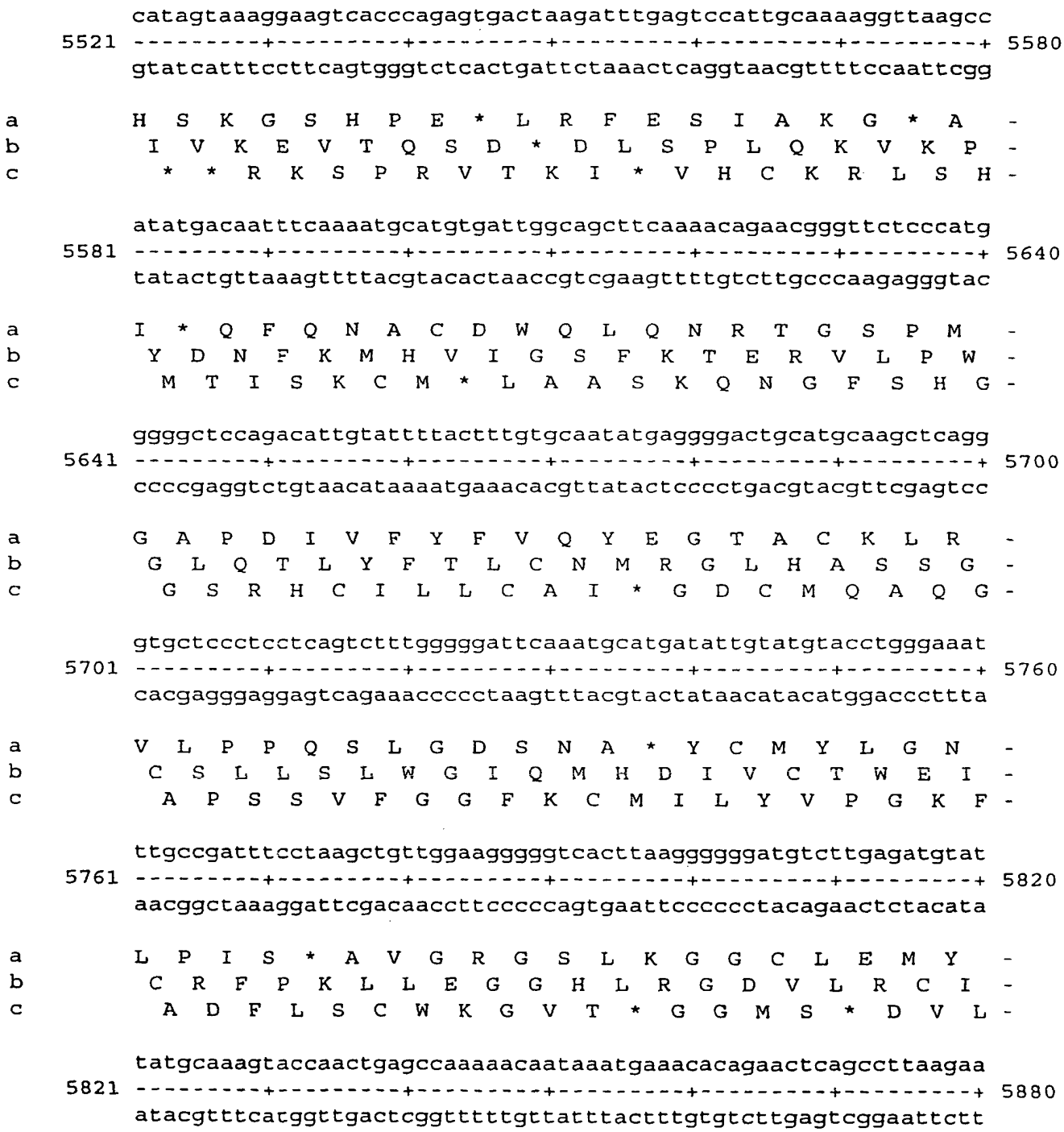
catagtaaaggaagtcacccagagtgactaagatttgagtccattgcaaaaggttaagcc
5521 ---------+---------+---------+---------+---------+---------+ 5580
gtatcatttccttcagtgggtctcactgattctaaactcaggtaacgttttccaattcgg
a H S K G S H P E * L R F E S I A K G * A -
b I V K E V T Q S D * D L S P L Q K V K P -
c * * R K S P R V T K I * V H C K R L S H -
atatgacaatttcaaaatgcatgtgattggcagcttcaaaacagaacgggttctcccatg
5581 ---------+---------+---------+---------+---------+---------+ 5640
tatactgttaaagttttacgtacactaaccgtcgaagttttgtcttgcccaagagggtac
a I * Q F Q N A C D W Q L Q N R T G S P M -
b Y D N F K M H V I G S F K T E R V L P W -
c M T I S K C M * L A A S K Q N G F S H G -
ggggctccagacattgtattttactttgtgcaatatgaggggactgcatgcaagctcagg
5641 ---------+---------+---------+---------+---------+---------+ 5700
ccccgaggtctgtaacataaaatgaaacacgttatactcccctgacgtacgttcgagtcc
a G A P D I V F Y F V Q Y E G T A C K L R -
b G L Q T L Y F T L C N M R G L H A S S G -
c G S R H C I L L C A I * G D C M Q A Q G -
gtgctccctcctcagtctttgggggattcaaatgcatgatattgtatgtacctgggaaat
5701 ---------+---------+---------+---------+---------+---------+ 5760
cacgagggaggagtcagaaacccccctaagtttacgtactataacatacatggacccttta
a V L P P Q S L G D S N A * Y C M Y L G N -
b C S L L S L W G I Q M H D I V C T W E I -
c A P S S V F G G F K C M I L Y V P G K F -
ttgccgatttcctaagctgttggaaggggtcacttaaggggggatgtcttgagatgtat
5761 ---------+---------+---------+---------+---------+---------+ 5820
aacggctaaaggattcgacaaccttcccccagtgaattcccccctacagaactctacata
a L P I S * A V G R G S L K G G C L E M Y -
b C R F P K L L E G G H L R G D V L R C I -
c A D F L S C W K G V T * G G M S * D V L -
tatgcaaagtaccaactgagccaaaaacaataaatgaaacacagaactcagccttaagaa
5821 ---------+---------+---------+---------+---------+---------+ 5880
atacgtttcatggttgactcggtttttgttatttactttgtgtcttgagtcggaattctt

Figure 8-19

```
a         Y  A  K  Y  Q  L  S  Q  K  Q  *  M  K  H  R  T  Q  P  *  E  -
b          M  Q  S  T  N  *  A  K  N  N  K  *  N  T  E  L  S  L  K  K  -
c           C  K  V  P  T  E  P  K  T  I  N  E  T  Q  N  S  A  L  R  K -

          agctatatatgaataattatgtttacctcactggtgcatttaaaatggacttttgttcat
     5881 ---------+---------+---------+---------+---------+---------+ 5940
          tcgatatatacttattaatacaaatggagtgaccacgtaaattttacctgaaaacaagta

a         S  Y  I  *  I  I  M  F  T  S  L  V  H  L  K  W  T  F  V  H  -
b          A  I  Y  E  *  L  C  L  P  H  W  C  I  *  N  G  L  L  F  M  -
c           L  Y  M  N  N  Y  V  Y  L  T  G  A  F  K  M  D  F  C  S  W -

          gggagaacctcgttgacatgcacagtttgcaatcttatgttgatcgatgttaaacgtcac
     5941 ---------+---------+---------+---------+---------+---------+ 6000
          ccctcttggagcaactgtacgtgtcaaacgttagaatacaactagctacaatttgcagtg

a         G  R  T  S  L  T  C  T  V  C  N  L  M  L  I  D  V  K  R  H  -
b          G  E  P  R  *  H  A  Q  F  A  I  L  C  *  S  M  L  N  V  T  -
c           E  N  L  V  D  M  H  S  L  Q  S  Y  V  D  R  C  *  T  S  Q -

          agcagtacttgctcaataaaggtcatattggaaacatagtcaattgctgagtcttatgtc
     6001 ---------+---------+---------+---------+---------+---------+ 6060
          tcgtcatgaacgagttatttccagtataacctttgtatcagttaacgactcagaatacag

a         S  S  T  C  S  I  K  V  I  L  E  T  *  S  I  A  E  S  Y  V  -
b          A  V  L  A  Q  *  R  S  Y  W  K  H  S  Q  L  L  S  L  M  S  -
c           Q  Y  L  L  N  K  G  H  I  G  N  I  V  N  C  *  V  L  C  H -

          atttctctttttctaatttttatttatttatttttatttagagatggggtcttgctatgt
     6061 ---------+---------+---------+---------+---------+---------+ 6120
          taaagagaaaaagattaaaaataaataaataaaaataaatctctaccccagaacgatacа

a         I  S  L  F  L  I  F  I  Y  L  F  L  F  R  D  G  V  L  L  C  -
b          F  L  F  F  *  F  L  F  I  Y  F  Y  L  E  M  G  S  C  Y  V  -
c           F  S  F  S  N  F  Y  L  F  I  F  I  *  R  W  G  L  A  M  W -

          ggcctcaagcagtcctcccacctcagccacccaaagtgctgggattacaggcatgagcca
     6121 ---------+---------+---------+---------+---------+---------+ 6180
          ccggagttcgtcaggagggtggagtcggtgggtttcacgaccctaatgtccgtactcggt

a         G  L  K  Q  S  S  H  L  S  H  P  K  C  W  D  Y  R  H  E  P  -
b          A  S  S  S  P  P  T  S  A  T  Q  S  A  G  I  T  G  M  S  H  -
c           P  Q  A  V  L  P  P  Q  P  P  K  V  L  G  L  Q  A  *  A  T -
```

Figure 8-20

```
        ccacgcccagcctgttatgccatttcaaagtgaaatctccactacctgaagcttgc
    6181 ---------+---------+---------+---------+---------+------ 6236
        ggtgcgggtcggacaatacggtaaagtttcactttagaggtgatggacttcgaacg

a         P  R  P  A  C  Y  A  I  S  K  *  N  L  H  Y  L  K  L    -
b          H  A  Q  P  V  M  P  F  Q  S  E  I  S  T  T  *  S  L   -
c           T  P  S  L  L  C  H  F  K  V  K  S  P  L  P  E  A  C  -
```

Enzymes that do cut:

NONE

Enzymes that do not cut:

NotI

Figure 9-1 Neurofilament H

(Linear) MAP of: hsnfh1.gcg check: 1349 from: 1 to: 1162

ID HSNFH1 standard; DNA; HUM; 1162 BP.
XX
AC X15306; X12501;
XX
NI g35028
XX . . .

With 1 enzymes: NOTI

October 31, 1996 14:30 ..

```
          ccactccggagtcctctgcccgcttcccgacctcgagggtctcctctgacgcgcagcgtc
        1 ---------+---------+---------+---------+---------+---------+ 60
          ggtgaggcctcaggagacgggcgaagggctggagctcccagaggagactgcgcgtcgcag

a         P  L  R  S  P  L  P  A  S  R  P  R  G  S  P  L  T  R  S  V  -
b          H  S  G  V  L  C  P  L  P  D  L  E  G  L  L  *  R  A  A  S  -
c           T  P  E  S  S  A  R  F  P  T  S  R  V  S  S  D  A  Q  R  R -

          gattccccttccctcctcggtcccctgccccgcccctctcactgcgcggagccggtcgcc
       61 ---------+---------+---------+---------+---------+---------+ 120
          ctaaggggaagggaggagccaggggacggggcggggagagtgacgcgcctcggccagcgg

a         D  S  P  S  L  L  G  P  L  P  R  P  S  H  C  A  E  P  V  A  -
b          I  P  L  P  S  S  V  P  C  P  A  P  L  T  A  R  S  R  S  P  -
c           F  P  F  P  P  R  S  P  A  P  P  L  S  L  R  G  A  G  R  R -

          gggggccgcaggggaggaggcggagaggcggggccctcctccccaccctctcactgcca
      121 ---------+---------+---------+---------+---------+---------+ 180
          ccccccggcgtcccctcctccgcctctccgccccgggaggaggggtgggagagtgacggt

a         G  G  P  Q  G  R  R  R  R  G  G  A  L  L  P  T  L  S  L  P  -
b          G  G  R  R  G  G  G  G  E  A  G  P  S  S  P  P  S  H  C  Q  -
c           G  A  A  G  E  E  A  E  R  R  G  P  P  P  H  P  L  T  A  K -

          aggggttggacccggccgcggcggctataaaagggccggcgccctggtcgtgccgcagtg
      181 ---------+---------+---------+---------+---------+---------+ 240
          tccccaacctgggccggcgccgccgatattttcccggccgcgggaccagcacggcgtcac

a         R  G  W  T  R  P  R  R  L  *  K  G  R  R  P  G  R  A  A  V  -
b          G  V  G  P  G  R  G  G  Y  K  R  A  G  A  L  V  V  P  Q  C  -
c           G  L  D  P  A  A  A  A  I  K  G  P  A  P  W  S  C  R  S  A -
```

Figure 9-2

```
     cctcccgccccgtcccggcctcgcgcacctgctcaggccatgatgagcttcggcggcgcg
241  ---------+---------+---------+---------+---------+---------+ 300
     ggagggcggggcagggccggagcgcgtggacgagtccggtactactcgaagccgccgcgc

a      P  P  A  P  S  R  P  R  A  P  A  Q  A  M  M  S  F  G  G  A  -
b       L  P  P  R  P  G  L  A  H  L  L  R  P  *  *  A  S  A  A  R  -
c        S  R  P  V  P  A  S  R  T  C  S  G  H  D  E  L  R  R  R  G -

     gacgcgctgctgggcgccccgttcgcgccgctgcatggcggcggcagcctccactacgcg
301  ---------+---------+---------+---------+---------+---------+ 360
     ctgcgcgacgacccgcggggcaagcgcggcgacgtaccgccgccgtcggaggtgatgcgc

a      D  A  L  L  G  A  P  F  A  P  L  H  G  G  G  S  L  H  Y  A  -
b       T  R  C  W  A  P  R  S  R  R  C  M  A  A  A  A  S  T  T  R  -
c        R  A  A  G  R  P  V  R  A  A  A  W  R  R  Q  P  P  L  R  A -

     ctagcccgaaagggtggcgcaggcgggacgcgctccgccgctggctcctccagcggcttc
361  ---------+---------+---------+---------+---------+---------+ 420
     gatcgggctttcccaccgcgtccgccctgcgcgaggcggcgaccgaggaggtcgccgaag

a      L  A  R  K  G  G  A  G  G  T  R  S  A  A  G  S  S  S  G  F  -
b       *  P  E  R  V  A  Q  A  G  R  A  P  P  L  A  P  P  A  A  S  -
c        S  P  K  G  W  R  R  R  D  A  L  R  R  W  L  L  Q  R  L  P -

     cactcgtggacacggacgtccgtgagctccgtgtccgcctcgcccagccgcttccgtggc
421  ---------+---------+---------+---------+---------+---------+ 480
     gtgagcacctgtgcctgcaggcactcgaggcacaggcggagcgggtcggcgaaggcaccg

a      H  S  W  T  R  T  S  V  S  S  V  S  A  S  P  S  R  F  R  G  -
b       T  R  G  H  G  R  P  *  A  P  C  P  P  R  P  A  A  S  V  A  -
c        L  V  D  T  D  V  R  E  L  R  V  R  L  A  Q  P  L  P  W  R -

     gcaggcgccgcctcaagcaccgactcgctggacacgctgagcaacgggccggagggctgc
481  ---------+---------+---------+---------+---------+---------+ 540
     cgtccgcggcggagttcgtggctgagcgacctgtgcgactcgttgcccggcctcccgacg

a      A  G  A  A  S  S  T  D  S  L  D  T  L  S  N  G  P  E  G  C  -
b       Q  A  P  P  Q  A  P  T  R  W  T  R  *  A  T  G  R  R  A  A  -
c        R  R  R  L  K  H  R  L  A  G  H  A  E  Q  R  A  G  G  L  H -

     atggtggcggtggccacctcacgcagtgagaaggagcagctgcaggcgctgaacgaccgc
541  ---------+---------+---------+---------+---------+---------+ 600
     taccaccgccaccggtggagtgcgtcactcttcctcgtcgacgtccgcgacttgctggcg
```

Figure 9-3

```
a         M  V  A  V  A  T  S  R  S  E  K  E  Q  L  Q  A  L  N  D  R  -
b          W  W  R  W  P  P  H  A  V  R  R  S  S  C  R  R  *  T  T  A  -
c           G  G  G  G  H  L  T  Q  *  E  G  A  A  A  G  A  E  R  P  L -

          ttcgccgggtacatcgacaaggtgcggcagctggaggcgcacaaccgcagcctggagggc
      601 ---------+---------+---------+---------+---------+---------+ 660
          aagcggcccatgtagctgttccacgccgtcgacctccgcgtgttggcgtcggacctcccg

a         F  A  G  Y  I  D  K  V  R  Q  L  E  A  H  N  R  S  L  E  G  -
b          S  P  G  T  S  T  R  C  G  S  W  R  R  T  T  A  A  W  R  A  -
c           R  R  V  H  R  Q  G  A  A  A  G  G  A  Q  P  Q  P  G  G  R -

          gaggctgcggcgctgcggcagcagcaggcgggccgctccgctatgggcgagctgtacgag
      661 ---------+---------+---------+---------+---------+---------+ 720
          ctccgacgccgcgacgccgtcgtcgtccgcccggcgaggcgatacccgctcgacatgctc

a         E  A  A  A  L  R  Q  Q  Q  A  G  R  S  A  M  G  E  L  Y  E  -
b          R  L  R  R  C  G  S  S  R  R  A  A  P  L  W  A  S  C  T  S  -
c           G  C  G  A  A  A  A  A  G  G  P  L  R  Y  G  R  A  V  R  A -

          cgcgaggtccgcgagatgcgcggcgcggtgctgcgcctgggcgcggcgcgcggtcagcta
      721 ---------+---------+---------+---------+---------+---------+ 780
          gcgctccaggcgctctacgcgccgcgccacgacgcggacccgcgccgcgcgccagtcgat

a         R  E  V  R  E  M  R  G  A  V  L  R  L  G  A  A  R  G  Q  L  -
b          A  R  S  A  R  C  A  A  R  C  C  A  W  A  R  R  A  V  S  Y  -
c           R  G  P  R  D  A  R  R  G  A  A  P  G  R  G  A  R  S  A  T -

          cgcctggagcaggagcacctgctcgaggacatcgcgcacgtgcgccagcgcctagacgac
      781 ---------+---------+---------+---------+---------+---------+ 840
          gcggacctcgtcctcgtggacgagctcctgtagcgcgtgcacgcggtcgcggatctgctg

a         R  L  E  Q  E  H  L  L  E  D  I  A  H  V  R  Q  R  L  D  D  -
b          A  W  S  R  S  T  C  S  R  T  S  R  T  C  A  S  A  *  T  T  -
c           P  G  A  G  A  P  A  R  G  H  R  A  R  A  P  A  P  R  R  R -

          gaggcccggcagcgagaggaggccgaggcggcggcccgcgcgctggcgcgcttcgcgcag
      841 ---------+---------+---------+---------+---------+---------+ 900
          ctccgggccgtcgctctcctccggctccgccgccgggcgcgcgaccgcgcgaagcgcgtc

a         E  A  R  Q  R  E  E  A  E  A  A  A  R  A  L  A  R  F  A  Q  -
b          R  P  G  S  E  R  R  P  R  R  R  P  A  R  W  R  A  S  R  R  -
c           G  P  A  A  R  G  G  R  G  G  G  P  R  A  G  A  L  R  A  G -
```

Figure 9-4

```
        gaggccgaggcggcgcgcgtggacctgcagaagaaggcgcaggcgctgcaggaggagtgc
    901 ---------+---------+---------+---------+---------+---------+ 960
        ctccggctccgccgcgcgcacctggacgtcttcttccgcgtccgcgacgtcctcctcacg

a       E  A  E  A  A  R  V  D  L  Q  K  K  A  Q  A  L  Q  E  E  C  -
b        R  P  R  R  R  A  W  T  C  R  R  R  R  R  C  R  R  S  A  -
c         G  R  G  G  A  R  G  P  A  E  E  G  A  G  A  A  G  G  V  R -

        ggctacctgcggcgccaccaccaggaagaggtgggcgagctgctcggccagatccagggc
    961 ---------+---------+---------+---------+---------+---------+ 1020
        ccgatggacgccgcggtggtggtccttctccacccgctcgacgagccggtctaggtcccg

a       G  Y  L  R  R  H  H  Q  E  E  V  G  E  L  L  G  Q  I  Q  G  -
b        A  T  C  G  A  T  T  R  K  R  W  A  S  C  S  A  R  S  R  A  -
c         L  P  A  A  P  P  P  G  R  G  G  R  A  A  R  P  D  P  G  L -

        tccggcgccgcgcaggcgcagatgcaggccgagacgcgcgacgccctgaagtgcgacgtg
   1021 ---------+---------+---------+---------+---------+---------+ 1080
        aggccgcggcgcgtccgcgtctacgtccggctctgcgcgctgcgggacttcacgctgcac

a       S  G  A  A  Q  A  Q  M  Q  A  E  T  R  D  A  L  K  C  D  V  -
b        P  A  P  R  R  R  R  C  R  P  R  R  A  T  P  *  S  A  T  *  -
c         R  R  R  A  G  A  D  A  G  R  D  A  R  R  P  E  V  R  R  D -

        acgtcggcgctgcgcgagattcgcgcgcagcttgaaggccacgcggtgcagagcacgctg
   1081 ---------+---------+---------+---------+---------+---------+ 1140
        tgcagccgcgacgcgctctaagcgcgcgtcgaacttccggtgcgccacgtctcgtgcgac

a       T  S  A  L  R  E  I  R  A  Q  L  E  G  H  A  V  Q  S  T  L  -
b        R  R  R  C  A  R  F  A  R  S  L  K  A  T  R  C  R  A  R  C  -
c         V  G  A  A  R  D  S  R  A  A  *  R  P  R  G  A  E  H  A  A -

        cagtccgaggagtggttccgag
   1141 ---------+---------+-- 1162
        gtcaggctcctcaccaaggctc

a       Q  S  E  E  W  F  R     -
b        S  P  R  S  G  S  E     -
c         V  R  G  V  V  P       -
```

Enzymes that do cut: NONE

Enzymes that do not cut: NotI

Figure 10-1 Presenilin I

(Linear) MAP of: hsu40379 check: 135 from: 1 to: 1392

RL;HSU40379 - Human presenilin I-463 (AD3-3) mRNA, complete cds.
ID HSU40379 standard; RNA; HUM; 1392 BP.
AC U40379;
NI g1244637
DT 05-APR-1996 (Rel. 47, Created)
DT 15-AUG-1996 (Rel. 48, Last updated, Version 3) . . .

With 1 enzymes: NOTI

```
        atgacagagttacctgcaccgttgtcctacttccagaatgcacagatgtctgaggacaac
      1 ---------+---------+---------+---------+---------+---------+ 60
        tactgtctcaatggacgtggcaacaggatgaaggtcttacgtgtctacagactcctgttg

a       M  T  E  L  P  A  P  L  S  Y  F  Q  N  A  Q  M  S  E  D  N  -
b        *  Q  S  Y  L  H  R  C  P  T  S  R  M  H  R  C  L  R  T  T -
c         D  R  V  T  C  T  V  V  L  L  P  E  C  T  D  V  *  G  Q  P -

        cacctgagcaatactaatgacaatagagaacggcaggagcacaacgacagacggagcctt
     61 ---------+---------+---------+---------+---------+---------+ 120
        gtggactcgttatgattactgttatctcttgccgtcctcgtgttgctgtctgcctcggaa

a       H  L  S  N  T  N  D  N  R  E  R  Q  E  H  N  D  R  R  S  L  -
b        T  *  A  I  L  M  T  I  E  N  G  R  S  T  T  T  D  G  A  L -
c         P  E  Q  Y  *  *  Q  *  R  T  A  G  A  Q  R  Q  T  E  P  W -

        ggccaccctgagccattatctaatggacgaccccagggtaactcccggcaggtggtggag
    121 ---------+---------+---------+---------+---------+---------+ 180
        ccggtgggactcggtaatagattacctgctggggtcccattgagggccgtccaccacctc

a       G  H  P  E  P  L  S  N  G  R  P  Q  G  N  S  R  Q  V  V  E  -
b        A  T  L  S  H  Y  L  M  D  D  P  R  V  T  P  G  R  W  W  S -
c         P  P  *  A  I  I  *  W  T  T  P  G  *  L  P  A  G  G  G  A -

        caagatgaggaagaagatgaggagctgacattgaaatatggcgccaagcatgtgatcatg
    181 ---------+---------+---------+---------+---------+---------+ 240
        gttctactccttcttctactcctcgactgtaactttataccgcggttcgtacactagtac

a       Q  D  E  E  E  D  E  E  L  T  L  K  Y  G  A  K  H  V  I  M  -
b        K  M  R  K  K  M  R  S  *  H  *  N  M  A  P  S  M  *  S  C -
c         R  *  G  R  R  *  G  A  D  I  E  I  W  R  Q  A  C  D  H  A -
```

Figure 10-2

```
        ctctttgtccctgtgactctctgcatggtggtggtcgtggctaccattaagtcagtcagc
    241 ---------+---------+---------+---------+---------+---------+ 300
        gagaaacagggacactgagagacgtaccaccaccagcaccgatggtaattcagtcagtcg

a           L  F  V  P  V  T  L  C  M  V  V  V  V  A  T  I  K  S  V  S  -
b            S  L  S  L  *  L  S  A  W  W  W  S  W  L  P  L  S  Q  S  A  -
c             L  C  P  C  D  S  L  H  G  G  G  R  G  Y  H  *  V  S  Q  L  -

        ttttatacccggaaggatgggcagctaatctataccccattcacagaagataccgagact
    301 ---------+---------+---------+---------+---------+---------+ 360
        aaaatatgggccttcctacccgtcgattagatatggggtaagtgtcttctatggctctga

a           F  Y  T  R  K  D  G  Q  L  I  Y  T  P  F  T  E  D  T  E  T  -
b            F  I  P  G  R  M  G  S  *  S  I  P  H  S  Q  K  I  P  R  L  -
c             L  Y  P  E  G  W  A  A  N  L  Y  P  I  H  R  R  Y  R  D  C  -

        gtgggccagagagccctgcactcaattctgaatgctgccatcatgatcagtgtcattgtt
    361 ---------+---------+---------+---------+---------+---------+ 420
        cacccggtctctcgggacgtgagttaagacttacgacggtagtactagtcacagtaacaa

a           V  G  Q  R  A  L  H  S  I  L  N  A  A  I  M  I  S  V  I  V  -
b            W  A  R  E  P  C  T  Q  F  *  M  L  P  S  *  S  V  S  L  L  -
c             G  P  E  S  P  A  L  N  S  E  C  C  H  H  D  Q  C  H  C  C  -

        gtcatgactatcctcctggtggttctgtataaatacaggtgctataaggtcatccatgcc
    421 ---------+---------+---------+---------+---------+---------+ 480
        cagtactgataggaggaccaccaagacatatttatgtccacgatattccagtaggtacgg

a           V  M  T  I  L  L  V  V  L  Y  K  Y  R  C  Y  K  V  I  H  A  -
b            S  *  L  S  S  W  W  F  C  I  N  T  G  A  I  R  S  S  M  P  -
c             H  D  Y  P  P  G  G  S  V  *  I  Q  V  L  *  G  H  P  C  L  -

        tggcttattatatcatctctattgttgctgttctttttttcattcatttacttgggggaa
    481 ---------+---------+---------+---------+---------+---------+ 540
        accgaataatatagtagagataacaacgacaagaaaaaaagtaagtaaatgaacccccctt

a           W  L  I  I  S  S  L  L  L  L  F  F  F  S  F  I  Y  L  G  E  -
b            G  L  L  Y  H  L  Y  C  C  C  S  F  F  H  S  F  T  W  G  K  -
c             A  Y  Y  I  I  S  I  V  A  V  L  F  F  I  H  L  L  G  G  S  -

        gtgtttaaaacctataacgttgctgtggactacattactgttgcactcctgatctggaat
    541 ---------+---------+---------+---------+---------+---------+ 600
        cacaaattttggatattgcaacgacacctgatgtaatgacaacgtgaggactagacctta
```

Figure 10-3

```
a        V  F  K  T  Y  N  V  A  V  D  Y  I  T  V  A  L  L  I  W  N  -
b         C  L  K  P  I  T  L  L  W  T  T  L  L  L  H  S  *  S  G  I  -
c          V  *  N  L  *  R  C  C  G  L  H  Y  C  C  T  P  D  L  E  F  -

         tttggtgtggtgggaatgatttccattcactggaaaggtccacttcgactccagcaggca
     601 ---------+---------+---------+---------+---------+---------+---------+ 660
         aaaccacaccacccttactaaaggtaagtgacctttccaggtgaagctgaggtcgtccgt

a        F  G  V  V  G  M  I  S  I  H  W  K  G  P  L  R  L  Q  Q  A  -
b         L  V  W  W  E  *  F  P  F  T  G  K  V  H  F  D  S  S  R  H  -
c          W  C  G  G  N  D  F  H  S  L  E  R  S  T  S  T  P  A  G  I  -

         tatctcattatgattagtgccctcatggccctggtgtttatcaagtacctccctgaatgg
     661 ---------+---------+---------+---------+---------+---------+---------+ 720
         atagagtaatactaatcacgggagtaccgggaccacaaatagttcatggagggacttacc

a        Y  L  I  M  I  S  A  L  M  A  L  V  F  I  K  Y  L  P  E  W  -
b         I  S  L  *  L  V  P  S  W  P  W  C  L  S  S  T  S  L  N  G  -
c          S  H  Y  D  *  C  P  H  G  P  G  V  Y  Q  V  P  P  *  M  D  -

         actgcgtggctcatcttggctgtgatttcagtatatgatttagtggctgttttgtgtccg
     721 ---------+---------+---------+---------+---------+---------+---------+ 780
         tgacgcaccgagtagaaccgacactaaagtcatatactaaatcaccgacaaaacacaggc

a        T  A  W  L  I  L  A  V  I  S  V  Y  D  L  V  A  V  L  C  P  -
b         L  R  G  S  S  W  L  *  F  Q  Y  M  I  *  W  L  F  C  V  R  -
c          C  V  A  H  L  G  C  D  F  S  I  *  F  S  G  C  F  V  S  E  -

         aaaggtccacttcgtatgctggttgaaacagcccaggagagaaatgaaacgcttttcca
     781 ---------+---------+---------+---------+---------+---------+---------+ 840
         tttccaggtgaagcatacgaccaactttgtcgggtcctctctttactttgcgaaaaaggt

a        K  G  P  L  R  M  L  V  E  T  A  Q  E  R  N  E  T  L  F  P  -
b         K  V  H  F  V  C  W  L  K  Q  P  R  R  E  M  K  R  F  F  Q  -
c          R  S  T  S  Y  A  G  *  N  S  P  G  E  K  *  N  A  F  S  S  -

         gctctcatttactcctcaacaatggtgtggttggtgaatatggcagaaggagacccggaa
     841 ---------+---------+---------+---------+---------+---------+---------+ 900
         cgagagtaaatgaggagttgttaccacaccaaccacttataccgtcttcctctgggcctt

a        A  L  I  Y  S  S  T  M  V  W  L  V  N  M  A  E  G  D  P  E  -
b         L  S  F  T  P  Q  Q  W  C  G  W  *  I  W  Q  K  E  T  R  K  -
c          S  H  L  L  L  N  N  G  V  V  G  E  Y  G  R  R  R  P  G  S  -
```

Figure 10-4

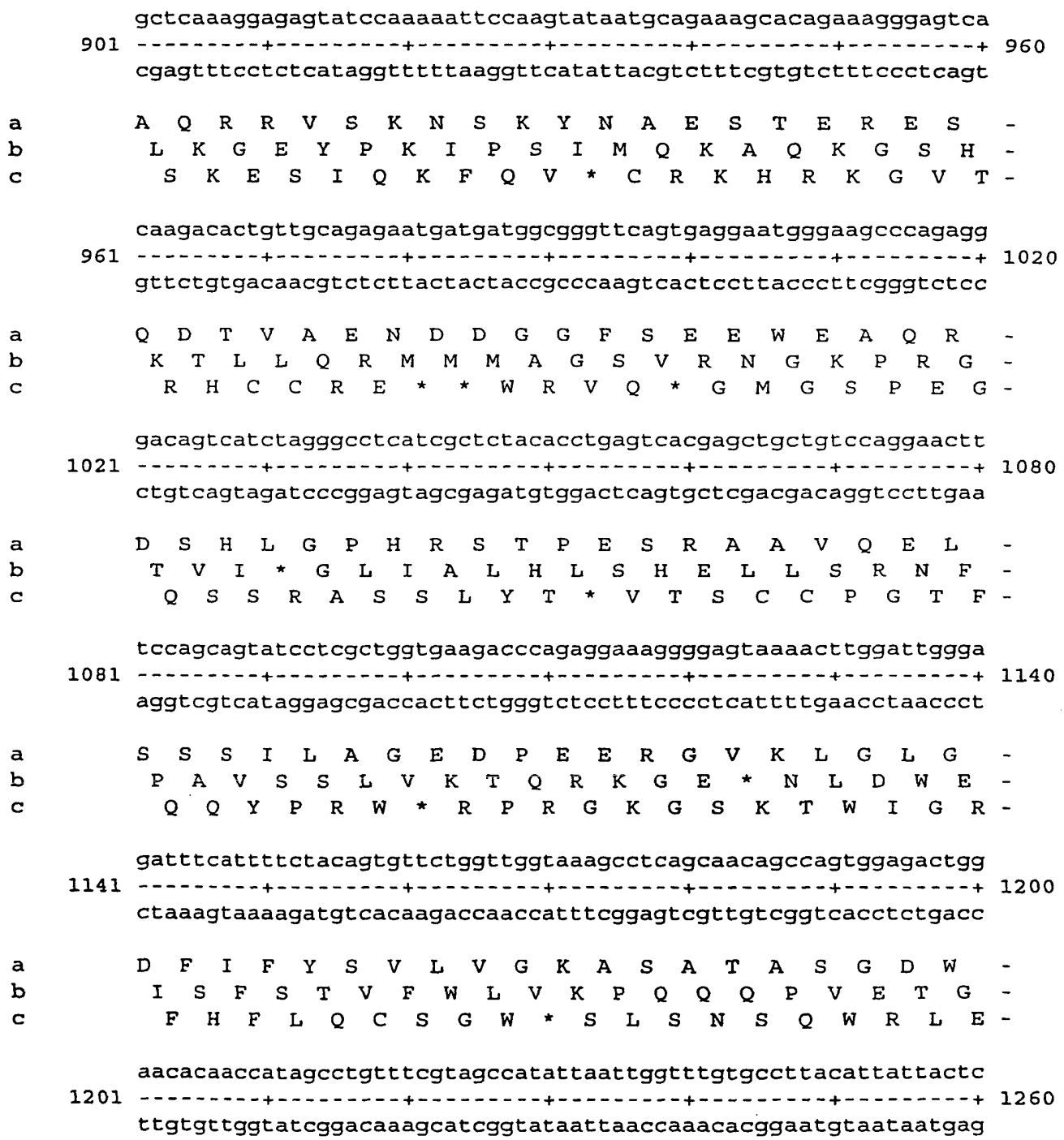
gctcaaaggagagtatccaaaaattccaagtataatgcagaaagcacagaaagggagtca
901 960
cgagtttcctctcataggtttttaaggttcatattacgtctttcgtgtctttccctcagt
a A Q R R V S K N S K Y N A E S T E R E S -
b L K G E Y P K I P S I M Q K A Q K G S H -
c S K E S I Q K F Q V * C R K H R K G V T -
caagacactgttgcagagaatgatgatggcgggttcagtgaggaatgggaagcccagagg
961 1020
gttctgtgacaacgtctcttactactaccgcccaagtcactccttacccttcgggtctcc
a Q D T V A E N D D G G F S E E W E A Q R -
b K T L L Q R M M M A G S V R N G K P R G -
c R H C C R E * * W R V Q * G M G S P E G -
gacagtcatctagggcctcatcgctctacacctgagtcacgagctgctgtccaggaactt
1021 1080
ctgtcagtagatcccggagtagcgagatgtggactcagtgctcgacgacaggtccttgaa
a D S H L G P H R S T P E S R A A V Q E L -
b T V I * G L I A L H L S H E L L S R N F -
c Q S S R A S S L Y T * V T S C C P G T F -
tccagcagtatcctcgctggtgaagacccagaggaaaggggagtaaaacttggattggga
1081 1140
aggtcgtcataggagcgaccacttctgggtctcctttcccctcattttgaacctaaccct
a S S S I L A G E D P E E R G V K L G L G -
b P A V S S L V K T Q R K G E * N L D W E -
c Q Q Y P R W * R P R G K G S K T W I G R -
gatttcattttctacagtgttctggttggtaaagcctcagcaacagccagtggagactgg
1141 1200
ctaaagtaaaagatgtcacaagaccaaccatttcggagtcgttgtcggtcacctctgacc
a D F I F Y S V L V G K A S A T A S G D W -
b I S F S T V F W L V K P Q Q Q P V E T G -
c F H F L Q C S G W * S L S N S Q W R L E -
aacacaaccatagcctgtttcgtagccatattaattggtttgtgccttacattattactc
1201 1260
ttgtgttggtatcggacaaagcatcggtataattaaccaaacacggaatgtaataatgag

Figure 10-5

```
a        N  T  T  I  A  C  F  V  A  I  L  I  G  L  C  L  T  L  L  L  -
b         T  Q  P  *  P  V  S  *  P  Y  *  L  V  C  A  L  H  Y  Y  S  -
c          H  N  H  S  L  F  R  S  H  I  N  W  F  V  P  Y  I  I  T  P -

         cttgccattttcaagaaagcattgccagctcttccaatctccatcacctttgggcttgtt
    1261 ---------+---------+---------+---------+---------+---------+ 1320
         gaacggtaaaagttctttcgtaacggtcgagaaggttagaggtagtggaaacccgaacaa

a        L  A  I  F  K  K  A  L  P  A  L  P  I  S  I  T  F  G  L  V  -
b         L  P  F  S  R  K  H  C  Q  L  F  Q  S  P  S  P  L  G  L  F  -
c          C  H  F  Q  E  S  I  A  S  S  S  N  L  H  H  L  W  A  C  F -

         ttctactttgccacagattatcttgtacagcctttatggaccaattagcattccatcaa
    1321 ---------+---------+---------+---------+---------+---------+ 1380
         aagatgaaacggtgtctaatagaacatgtcggaaaatacctggttaatcgtaaggtagtt

a        F  Y  F  A  T  D  Y  L  V  Q  P  F  M  D  Q  L  A  F  H  Q  -
b         S  T  L  P  Q  I  I  L  Y  S  L  L  W  T  N  *  H  S  I  N  -
c          L  L  C  H  R  L  S  C  T  A  F  Y  G  P  I  S  I  P  S  I -

         ttttatatctag
    1381 ----------+-- 1392
         aaaatatagatc

a        F  Y  I  *  -
b         F  I  S  -
c          L  Y  L  -
```

Enzymes that do cut:

NONE

Enzymes that do not cut:

NotI

Figure 11-1 Presenilin II

(Linear) MAP of: hsstm2r check: 9487 from: 1 to: 2236

RL;HSSTM2R - Homo sapiens (clone F-T03796) STM-2 mRNA, complete cds.
ID HSSTM2R standard; RNA; HUM; 2236 BP.
AC L43964;
NI g951202
DT 24-AUG-1995 (Rel. 44, Created)
DT 17-FEB-1997 (Rel. 50, Last updated, Version 2) . . .

With 1 enzymes: NOTI

```
          cgagcggcggcggagcaggcatttccagcagtgaggagacagccagaagcaagctattgg
        1 ---------+---------+---------+---------+---------+---------+ 60
          gctcgccgccgcctcgtccgtaaaggtcgtcactcctctgtcggtcttcgttcgataacc

a         R  A  A  A  E  Q  A  F  P  A  V  R  R  Q  P  E  A  S  Y  W  -
b          E  R  R  R  S  R  H  F  Q  Q  *  G  D  S  Q  K  Q  A  I  G  -
c           S  G  G  G  A  G  I  S  S  S  E  E  T  A  R  S  K  L  L  E -

          agctgaaggaacctgagacagaagctagtccccctctgaattttactgatgaagaaact
       61 ---------+---------+---------+---------+---------+---------+ 120
          tcgacttccttggactctgtcttcgatcaggggggagacttaaaatgactacttctttga

a         S  *  R  N  L  R  Q  K  L  V  P  P  L  N  F  T  D  E  E  T  -
b          A  E  G  T  *  D  R  S  *  S  P  L  *  I  L  L  M  K  K  L  -
c           L  K  E  P  E  T  E  A  S  P  P  S  E  F  Y  *  *  R  N  * -

          gaggccacagagctaaagtgacttttcccaaggtcgcccagcgaggacgtgggacttctc
      121 ---------+---------+---------+---------+---------+---------+ 180
          ctccggtgtctcgatttcactgaaaagggttccagcgggtcgctcctgcaccctgaagag

a         E  A  T  E  L  K  *  L  F  P  R  S  P  S  E  D  V  G  L  L  -
b          R  P  Q  S  *  S  D  F  S  Q  G  R  P  A  R  T  W  D  F  S  -
c           G  H  R  A  K  V  T  F  P  K  V  A  Q  R  G  R  G  T  S  Q -

          agacgtcaggagagtgatgtgagggagctgtgtgaccatagaaagtgacgtgttaaaaac
      181 ---------+---------+---------+---------+---------+---------+ 240
          tctgcagtcctctcactacactccctcgacacactggtatctttcactgcacaatttttg

a         R  R  Q  E  S  D  V  R  E  L  C  D  H  R  K  *  R  V  K  N  -
b          D  V  R  R  V  M  *  G  S  C  V  T  I  E  S  D  V  L  K  T  -
c           T  S  G  E  *  C  E  G  A  V  *  P  *  K  V  T  C  *  K  P -
```

Figure 11-2

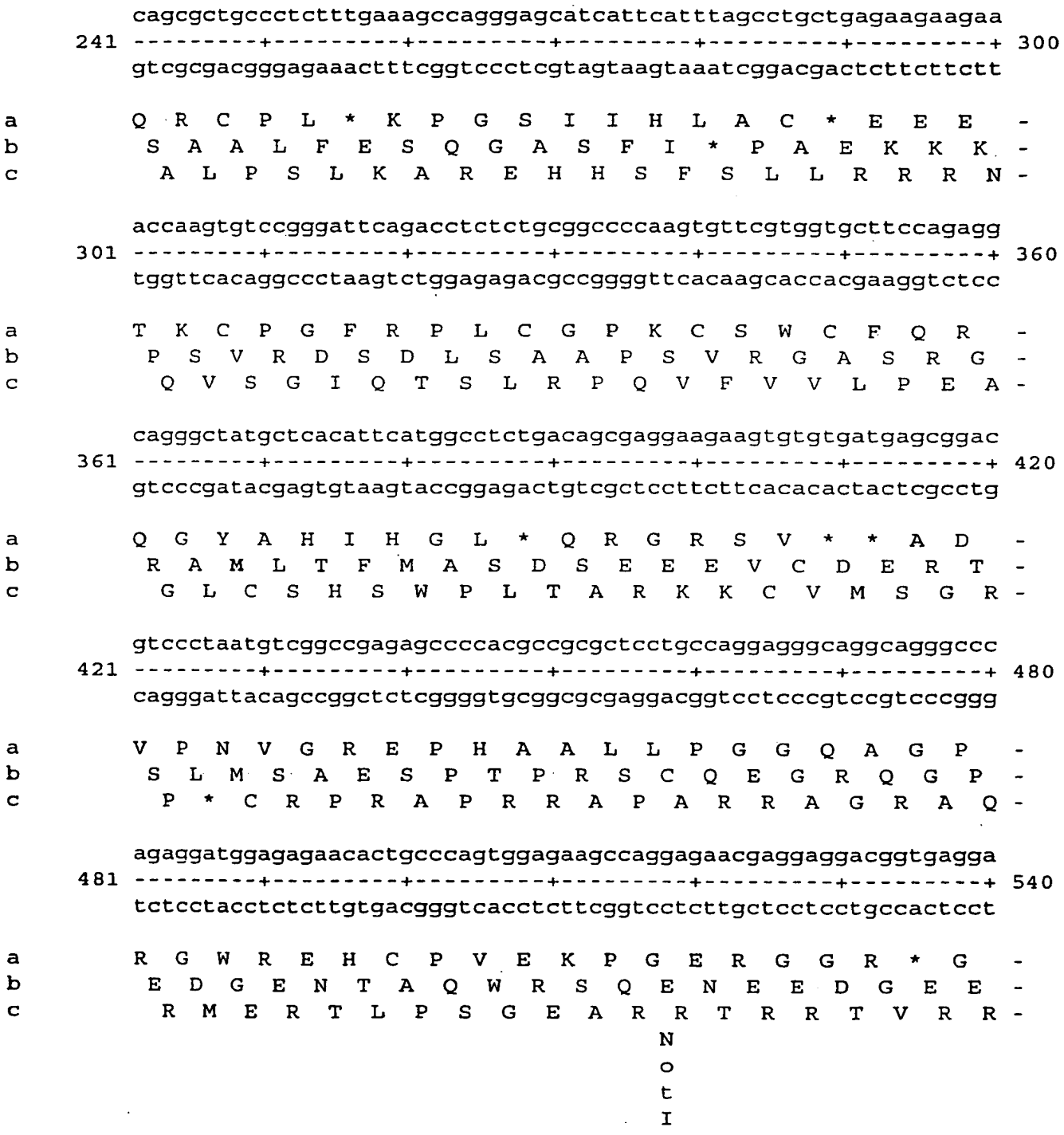
cagcgctgccctctttgaaagccagggagcatcattcatttagcctgctgagaagaagaa
241 ---------+---------+---------+---------+---------+---------+ 300
gtcgcgacgggagaaactttcggtccctcgtagtaagtaaatcggacgactcttcttctt
a Q R C P L * K P G S I I H L A C * E E E -
b S A A L F E S Q G A S F I * P A E K K K -
c A L P S L K A R E H H S F S L L R R R N -
accaagtgtccgggattcagacctctctgcggccccaagtgttcgtggtgcttccagagg
301 ---------+---------+---------+---------+---------+---------+ 360
tggttcacaggccctaagtctggagagacgccggggttcacaagcaccacgaaggtctcc
a T K C P G F R P L C G P K C S W C F Q R -
b P S V R D S D L S A A P S V R G A S R G -
c Q V S G I Q T S L R P Q V F V V L P E A -
cagggctatgctcacattcatggcctctgacagcgaggaagaagtgtgtgatgagcggac
361 ---------+---------+---------+---------+---------+---------+ 420
gtcccgatacgagtgtaagtaccggagactgtcgctccttcttcacacactactcgcctg
a Q G Y A H I H G L * Q R G R S V * * A D -
b R A M L T F M A S D S E E E V C D E R T -
c G L C S H S W P L T A R K K C V M S G R -
gtccctaatgtcggccgagagccccacgccgcgctcctgccaggagggcaggcagggccc
421 ---------+---------+---------+---------+---------+---------+ 480
cagggattacagccggctctcggggtgcggcgcgaggacggtcctcccgtccgtcccggg
a V P N V G R E P H A A L L P G G Q A G P -
b S L M S A E S P T P R S C Q E G R Q G P -
c P * C R P R A P R R A P A R R A G R A Q -
agaggatggagagaacactgcccagtggagaagccaggagaacgaggaggacggtgagga
481 ---------+---------+---------+---------+---------+---------+ 540
tctcctacctctcttgtgacgggtcacctcttcggtcctcttgctcctcctgccactcct
a R G W R E H C P V E K P G E R G G R * G -
b E D G E N T A Q W R S Q E N E E D G E E -
c R M E R T L P S G E A R R T R R T V R R -
N
o
t
I

Figure 11-3

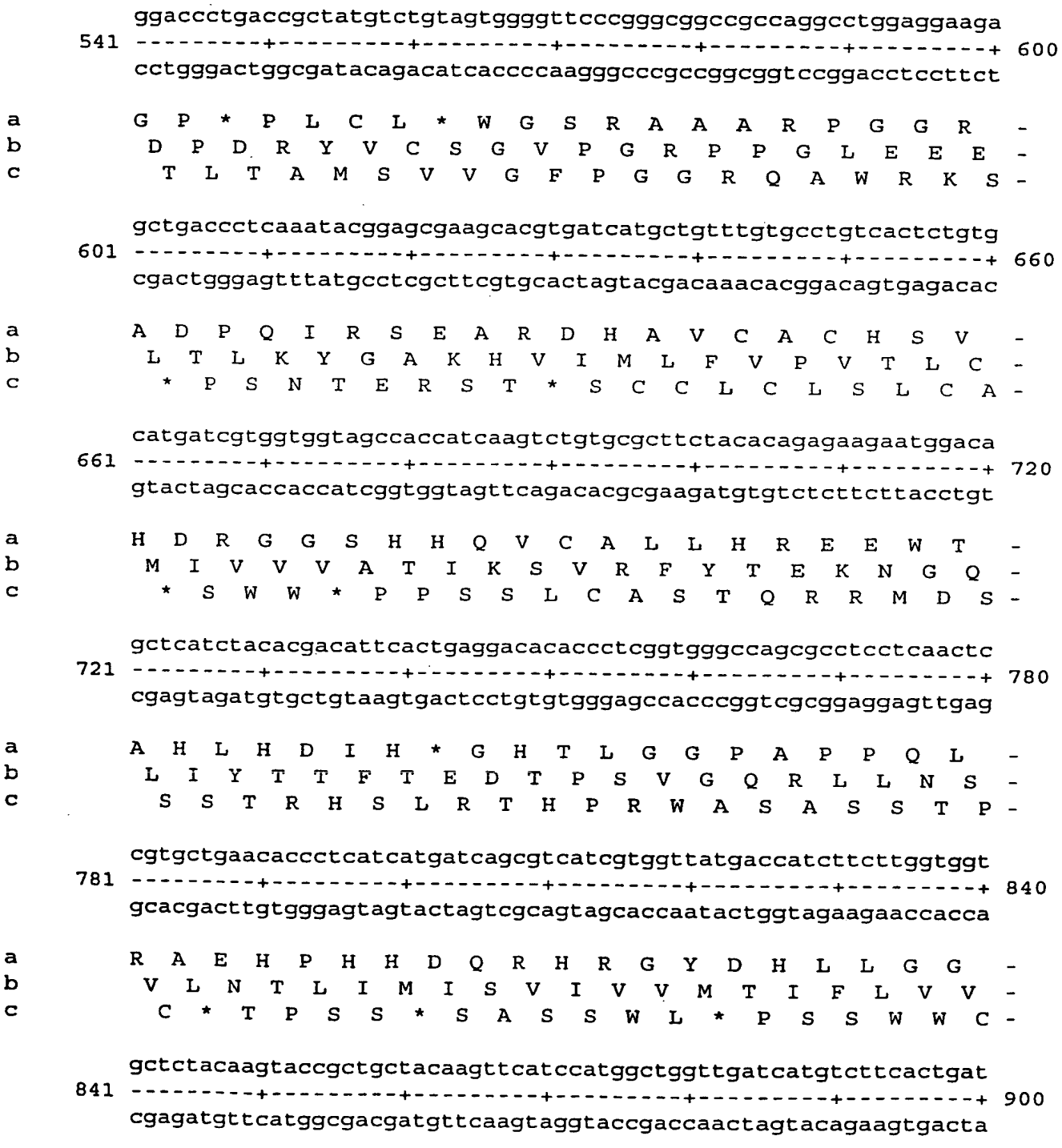
ggaccctgaccgctatgtctgtagtggggttcccgggcggccgccaggcctggaggaaga
541 ---------+---------+---------+---------+---------+---------+ 600
cctgggactggcgatacagacatcaccccaagggcccgccggcggtccggacctccttct
a G P * P L C L * W G S R A A A R P G G R -
b D P D R Y V C S G V P G R P P G L E E E -
c T L T A M S V V G F P G G R Q A W R K S -
gctgaccctcaaatacggagcgaagcacgtgatcatgctgtttgtgcctgtcactctgtg
601 ---------+---------+---------+---------+---------+---------+ 660
cgactgggagtttatgcctcgcttcgtgcactagtacgacaaacacggacagtgagacac
a A D P Q I R S E A R D H A V C A C H S V -
b L T L K Y G A K H V I M L F V P V T L C -
c * P S N T E R S T * S C C L C L S L C A -
catgatcgtggtggtagccaccatcaagtctgtgcgcttctacacagagaagaatggaca
661 ---------+---------+---------+---------+---------+---------+ 720
gtactagcaccaccatcggtggtagttcagacacgcgaagatgtgtctcttcttacctgt
a H D R G G S H H Q V C A L L H R E E W T -
b M I V V V A T I K S V R F Y T E K N G Q -
c * S W W * P P S S L C A S T Q R R M D S -
gctcatctacacgacattcactgaggacacaccctcggtgggccagcgcctcctcaactc
721 ---------+---------+---------+---------+---------+---------+ 780
cgagtagatgtgctgtaagtgactcctgtgtgggagccacccggtcgcggaggagttgag
a A H L H D I H * G H T L G G P A P P Q L -
b L I Y T T F T E D T P S V G Q R L L N S -
c S S T R H S L R T H P R W A S A S S T P -
cgtgctgaacaccctcatcatgatcagcgtcatcgtggttatgaccatcttcttggtggt
781 ---------+---------+---------+---------+---------+---------+ 840
gcacgacttgtgggagtagtactagtcgcagtagcaccaatactggtagaagaaccacca
a R A E H P H H D Q R H R G Y D H L L G G -
b V L N T L I M I S V I V V M T I F L V V -
c C * T P S S * S A S S W L * P S S W W C -
gctctacaagtaccgctgctacaagttcatccatggctggttgatcatgtcttcactgat
841 ---------+---------+---------+---------+---------+---------+ 900
cgagatgttcatggcgacgatgttcaagtaggtaccgaccaactagtacagaagtgacta

Figure 11-4

```
a        A  L  Q  V  P  L  L  Q  V  H  P  W  L  V  D  H  V  F  T  D  -
b         L  Y  K  Y  R  C  Y  K  F  I  H  G  W  L  I  M  S  S  L  M  -
c          S  T  S  T  A  A  T  S  S  S  M  A  G  *  S  C  L  H  *  C -

         gctgctgttcctcttcacctatatctaccttggggaagtgctcaagacctacaatgtggc
    901  ---------+---------+---------+---------+---------+---------+  960
         cgacgacaaggagaagtggatatagatggaacccttcacgagttctggatgttacaccg

a        A  A  V  P  L  H  L  Y  L  P  W  G  S  A  Q  D  L  Q  C  G  -
b         L  L  F  L  F  T  Y  I  Y  L  G  E  V  L  K  T  Y  N  V  A  -
c          C  C  S  S  S  P  I  S  T  L  G  K  C  S  R  P  T  M  W  P -

         catggactaccccaccctcttgctgactgtctggaacttcggggcagtgggcatggtgtg
    961  ---------+---------+---------+---------+---------+---------+ 1020
         gtacctgatggggtgggagaacgactgacagaccttgaagccccgtcacccgtaccacac

a        H  G  L  P  H  P  L  A  D  C  L  E  L  R  G  S  G  H  G  V  -
b         M  D  Y  P  T  L  L  L  T  V  W  N  F  G  A  V  G  M  V  C  -
c          W  T  T  P  P  S  C  *  L  S  G  T  S  G  Q  W  A  W  C  A -

         catccactggaagggccctctggtgctgcagcaggcctacctcatcatgatcagtgcgct
   1021  ---------+---------+---------+---------+---------+---------+ 1080
         gtaggtgaccttcccgggagaccacgacgtcgtccggatggagtagtactagtcacgcga

a        H  P  L  E  G  P  S  G  A  A  A  G  L  P  H  H  D  Q  C  A  -
b         I  H  W  K  G  P  L  V  L  Q  Q  A  Y  L  I  M  I  S  A  L  -
c          S  T  G  R  A  L  W  C  C  S  R  P  T  S  S  *  S  V  R  S -

         catggccctagtgttcatcaagtacctcccagagtggtccgcgtgggtcatcctgggcgc
   1081  ---------+---------+---------+---------+---------+---------+ 1140
         gtaccgggatcacaagtagttcatggagggtctcaccaggcgcacccagtaggacccgcg

a        H  G  P  S  V  H  Q  V  P  P  R  V  V  R  V  G  H  P  G  R  -
b         M  A  L  V  F  I  K  Y  L  P  E  W  S  A  W  V  I  L  G  A  -
c          W  P  *  C  S  S  S  T  S  Q  S  G  P  R  G  S  S  W  A  P -

         catctctgtgtatgatctcgtggctgtgctgtgtcccaaagggcctctgagaatgctggt
   1141  ---------+---------+---------+---------+---------+---------+ 1200
         gtagagacacatactagagcaccgacacgacacagggtttcccggagactcttacgacca

a        H  L  C  V  *  S  R  G  C  A  V  S  Q  R  A  S  E  N  A  G  -
b         I  S  V  Y  D  L  V  A  V  L  C  P  K  G  P  L  R  M  L  V  -
c          S  L  C  M  I  S  W  L  C  C  V  P  K  G  L  *  E  C  W  * -
```

Figure 11-5

```
         agaaactgcccaggagagaaatgagcccatattccctgccctgatatactcatctgccat
    1201 ---------+---------+---------+---------+---------+---------+ 1260
         tctttgacgggtcctctctttactcgggtataagggacgggactatatgagtagacggta

a        R  N  C  P  G  E  K  *  A  H  I  P  C  P  D  I  L  I  C  H  -
b         E  T  A  Q  E  R  N  E  P  I  F  P  A  L  I  Y  S  S  A  M  -
c          K  L  P  R  R  E  M  S  P  Y  S  L  P  *  Y  T  H  L  P  W -

         ggtgtggacggttggcatggcgaagctggacccctcctctcagggtgccctccagctccc
    1261 ---------+---------+---------+---------+---------+---------+ 1320
         ccacacctgccaaccgtaccgcttcgacctggggaggagagtcccacgggaggtcgaggg

a        G  V  D  G  W  H  G  E  A  G  P  L  L  S  G  C  P  P  A  P  -
b         V  W  T  V  G  M  A  K  L  D  P  S  S  Q  G  A  L  Q  L  P  -
c          C  G  R  L  A  W  R  S  W  T  P  P  L  R  V  P  S  S  S  P -

         ctacgacccggagatggaagaagactcctatgacagttttggggagccttcataccccga
    1321 ---------+---------+---------+---------+---------+---------+ 1380
         gatgctgggcctctaccttcttctgaggatactgtcaaaacccctcggaagtatggggct

a        L  R  P  G  D  G  R  R  L  L  *  Q  F  W  G  A  F  I  P  R  -
b         Y  D  P  E  M  E  E  D  S  Y  D  S  F  G  E  P  S  Y  P  E  -
c          T  T  R  R  W  K  K  T  P  M  T  V  L  G  S  L  H  T  P  K -

         agtctttgagcctcccttgactggctacccaggggaggagctggaggaagaggaggaaag
    1381 ---------+---------+---------+---------+---------+---------+ 1440
         tcagaaactcggagggaactgaccgatgggtcccctcctcgacctccttctcctcctttc

a        S  L  *  A  S  L  D  W  L  P  R  G  G  A  G  G  R  G  G  K  -
b         V  F  E  P  P  L  T  G  Y  P  G  E  E  L  E  E  E  E  E  R  -
c          S  L  S  L  P  *  L  A  T  Q  G  R  S  W  R  K  R  R  K  G -

         gggcgtgaagcttggcctcggggacttcatcttctacagtgtgctggtgggcaaggcggc
    1441 ---------+---------+---------+---------+---------+---------+ 1500
         cccgcacttcgaaccggagcccctgaagtagaagatgtcacacgaccacccgttccgccg

a        G  R  E  A  W  P  R  G  L  H  L  L  Q  C  A  G  G  Q  G  G  -
b         G  V  K  L  G  L  G  D  F  I  F  Y  S  V  L  V  G  K  A  A  -
c          A  *  S  L  A  S  G  T  S  S  S  T  V  C  W  W  A  R  R  L -

         tgccacgggcagcggggactggaataccacgctggcctgcttcgtggccatcctcattgg
    1501 ---------+---------+---------+---------+---------+---------+ 1560
         acggtgcccgtcgcccctgaccttatggtgcgaccggacgaagcaccggtaggagtaacc
```

Figure 11-6

```
a        C  H  G  Q  R  G  L  E  Y  H  A  G  L  L  R  G  H  P  H  W  -
b         A  T  G  S  G  D  W  N  T  T  L  A  C  F  V  A  I  L  I  G  -
c          P  R  A  A  G  T  G  I  P  R  W  P  A  S  W  P  S  S  L  A -

         cttgtgtctgaccctcctgctgcttgctgtgttcaagaaggcgctgcccgccctccccat
    1561 ---------+---------+---------+---------+---------+---------+---------+ 1620
         gaacacagactgggaggacgacgaacgacacaagttcttccgcgacgggcgggaggggta

a        L  V  S  D  P  P  A  A  C  C  V  Q  E  G  A  A  R  P  P  H  -
b         L  C  L  T  L  L  L  L  A  V  F  K  K  A  L  P  A  L  P  I  -
c          C  V  *  P  S  C  C  L  L  C  S  R  R  R  C  P  P  S  P  S -

         ctccatcacgttcgggctcatcttttacttctccacggacaacctggtgcggccgttcat
    1621 ---------+---------+---------+---------+---------+---------+---------+ 1680
         gaggtagtgcaagcccgagtagaaaatgaagaggtgcctgttggaccacgccggcaagta

a        L  H  H  V  R  A  H  L  L  L  L  H  G  Q  P  G  A  A  V  H  -
b         S  I  T  F  G  L  I  F  Y  F  S  T  D  N  L  V  R  P  F  M  -
c          P  S  R  S  G  S  S  F  T  S  P  R  T  T  W  C  G  R  S  W -

         ggacaccctggcctcccatcagctctacatctgagggacatggtgtgccacaggctgcaa
    1681 ---------+---------+---------+---------+---------+---------+---------+ 1740
         cctgtgggaccggagggtagtcgagatgtagactccctgtaccacacggtgtccgacgtt

a        G  H  P  G  L  P  S  A  L  H  L  R  D  M  V  C  H  R  L  Q  -
b         D  T  L  A  S  H  Q  L  Y  I  *  G  T  W  C  A  T  G  C  K  -
c          T  P  W  P  P  I  S  S  T  S  E  G  H  G  V  P  Q  A  A  S -

         gctgcagggaattttcattggatgcagttgtatagttttacactctagtgccatatattt
    1741 ---------+---------+---------+---------+---------+---------+---------+ 1800
         cgacgtcccttaaaagtaacctacgtcaacatatcaaaatgtgagatcacggtatataaa

a        A  A  G  N  F  H  W  M  Q  L  Y  S  F  T  L  *  C  H  I  F  -
b         L  Q  G  I  F  I  G  C  S  C  I  V  L  H  S  S  A  I  Y  F  -
c          C  R  E  F  S  L  D  A  V  V  *  F  Y  T  L  V  P  Y  I  F -

         ttaagacttttctttccttaaaaaataaagtacgtgtttacttggtgaggaggaggcaga
    1801 ---------+---------+---------+---------+---------+---------+---------+ 1860
         aattctgaaaagaaaggaatttttatttcatgcacaaatgaaccactcctcctccgtct

a        L  R  L  F  F  P  *  K  I  K  Y  V  F  T  W  *  G  G  G  R  -
b         *  D  F  S  F  L  K  K  *  S  T  C  L  L  G  E  E  E  A  E  -
c          K  T  F  L  S  L  K  N  K  V  R  V  Y  L  V  R  R  R  Q  N -
```

Figure 11-7

```
          accagctctttggtgccagctgtttcatcaccagactttggctcccgctttggggagcgc
     1861 ---------+---------+---------+---------+---------+---------+ 1920
          tggtcgagaaaccacggtcgacaaagtagtggtctgaaaccgagggcgaaacccctcgcg

a         T  S  S  L  V  P  A  V  S  S  P  D  F  G  S  R  F  G  E  R  -
b          P  A  L  W  C  Q  L  F  H  H  Q  T  L  A  P  A  L  G  S  A  -
c           Q  L  F  G  A  S  C  F  I  T  R  L  W  L  P  L  W  G  A  P -

          ctcgcttcacggacaggaagcacagcaggtttatccagatgaactgagaaggtcagatta
     1921 ---------+---------+---------+---------+---------+---------+ 1980
          gagcgaagtgcctgtccttcgtgtcgtccaaataggtctacttgactcttccagtctaat

a         L  A  S  R  T  G  S  T  A  G  L  S  R  *  T  E  K  V  R  L  -
b          S  L  H  G  Q  E  A  Q  Q  V  Y  P  D  E  L  R  R  S  D  *  -
c           R  F  T  D  R  K  H  S  R  F  I  Q  M  N  *  E  G  Q  I  R -

          gggcggggagaagagcatccggcatgagggctgagatgcgcaaagagtgtgctcgggagt
     1981 ---------+---------+---------+---------+---------+---------+ 2040
          cccgcccctcttctcgtaggccgtactcccgactctacgcgtttctcacacgagccctca

a         G  R  G  E  E  H  P  A  *  G  L  R  C  A  K  S  V  L  G  S  -
b          G  G  E  K  S  I  R  H  E  G  *  D  A  Q  R  V  C  S  G  V  -
c           A  G  R  R  A  S  G  M  R  A  E  M  R  K  E  C  A  R  E  W -

          ggcccctggcacctgggtgctctggctggagaggaaaagccagttccctacgaggagtgt
     2041 ---------+---------+---------+---------+---------+---------+ 2100
          ccggggaccgtggacccacgagaccgacctctccttttcggtcaagggatgctcctcaca

a         G  P  W  H  L  G  A  L  A  G  E  E  K  P  V  P  Y  E  E  C  -
b          A  P  G  T  W  V  L  W  L  E  R  K  S  Q  F  P  T  R  S  V  -
c           P  L  A  P  G  C  S  G  W  R  G  K  A  S  S  L  R  G  V  F -

          tcccaatgctttgtccatgatgtccttgttattttattgcctttagaaactgagtcctgt
     2101 ---------+---------+---------+---------+---------+---------+ 2160
          agggttacgaaacaggtactacaggaacaataaaataacggaaatctttgactcaggaca

a         S  Q  C  F  V  H  D  V  L  V  I  L  L  P  L  E  T  E  S  C  -
b          P  N  A  L  S  M  M  S  L  L  F  Y  C  L  *  K  L  S  P  V  -
c           P  M  L  C  P  *  C  P  C  Y  F  I  A  F  R  N  *  V  L  F -

          tcttgttacggcagtcacactgctgggaagtggcttaatagtaatatcaataaatagatg
     2161 ---------+---------+---------+---------+---------+---------+ 2220
          agaacaatgccgtcagtgtgacgacccttcaccgaattatcattatagttatttatctac
```

Figure 11-8

```
a        S  C  Y  G  S  H  T  A  G  K  W  L  N  S  N  I  N  K  *  M  -
b         L  V  T  A  V  T  L  L  G  S  G  L  I  V  I  S  I  N  R  *  -
c          L  L  R  Q  S  H  C  W  E  V  A  *  *  *  Y  Q  *  I  D  E -

         agtcctgttagaaaaa
    2221 ---------+------ 2236
         tcaggacaatcttttt

a        S  P  V  R  K    -
b         V  L  L  E  K   -
c          S  C  *  K     -
```

Enzymes that do cut:

NotI

Enzymes that do not cut:

NONE

Figure 12-1 Big Tau (exon 4A)

(Linear) MAP of: af047858 check: 5416 from: 1 to: 954

DL;AF047858 - Homo sapiens microtubule-associated protein tau (tau) gene, exon
ID AF047858 standard; DNA; HUM; 954 BP.
AC AF047858; M93652;
NI g2898166
DT 23-FEB-1998 (Rel. 54, Created)
DT 26-FEB-1998 (Rel. 54, Last updated, Version 4) . . .

With 1 enzymes: NOTI ***start exon 4A**

```
          gactgggccgagaagggtccggcctttccgaagcccgccaccactgcgtatctccacaca
        1 ---------+---------+---------+---------+---------+---------+ 60
          ctgacccggctcttcccaggccggaaaggcttcgggcggtggtgacgcatagaggtgtgt

a        *D  W  A  E  K  G  P  A  F  P  K  P  A  T  T  A  Y  L  H  T  -
b         T  G  P  R  R  V  R  P  F  R  S  P  P  P  L  R  I  S  T  Q  -
c          L  G  R  E  G  S  G  L  S  E  A  R  H  H  C  V  S  P  H  R -

          gagcctgaaagtggtaaggtggtccaggaaggcttcctccgagagccaggccccccaggt
       61 ---------+---------+---------+---------+---------+---------+ 120
          ctcggactttcaccattccaccaggtccttccgaaggaggctctcggtccgggggggtcca

a         E  P  E  S  G  K  V  V  Q  E  G  F  L  R  E  P  G  P  P  G  -
b         S  L  K  V  V  R  W  S  R  K  A  S  S  E  S  Q  A  P  Q  V  -
c          A  *  K  W  *  G  G  P  G  R  L  P  P  R  A  R  P  P  R  S -

          ctgagccaccagctcatgtccggcatgcctggggctcccctcctgcctgagggccccaga
      121 ---------+---------+---------+---------+---------+---------+ 180
          gactcggtggtcgagtacaggccgtacggaccccgaggggaggacggactcccggggtct

a         L  S  H  Q  L  M  S  G  M  P  G  A  P  L  L  P  E  G  P  R  -
b          *  A  T  S  S  C  P  A  C  L  G  L  P  S  C  L  R  A  P  E  -
c           E  P  P  A  H  V  R  H  A  W  G  S  P  P  A  *  G  P  Q  R -
                                                        N
                                                        o
                                                        t
                                                        I

          gaggccacacgccaaccttcggggacaggacctgaggacacagagggcggccgccacgcc
      181 ---------+---------+---------+---------+---------+---------+ 240
          ctccggtgtgcggttggaagcccctgtcctggactcctgtgtctcccgccggcggtgcgg
```

Figure 12-2

```
a        E  A  T  R  Q  P  S  G  T  G  P  E  D  T  E  G  G  R  H  A  -
b         R  P  H  A  N  L  R  G  Q  D  L  R  T  Q  R  A  A  A  T  P  -
c          G  H  T  P  T  F  G  D  R  T  *  G  H  R  G  R  P  P  R  P -

         cctgagctgctcaagcaccagcttctaggagacctgcaccaggaggggccgccgctgaag
     241 ---------+---------+---------+---------+---------+---------+ 300
         ggactcgacgagttcgtggtcgaagatcctctggacgtggtcctccccggcggcgacttc

a        P  E  L  L  K  H  Q  L  L  G  D  L  H  Q  E  G  P  P  L  K  -
b         L  S  C  S  S  T  S  F  *  E  T  C  T  R  R  G  R  R  *  R  -
c          *  A  A  Q  A  P  A  S  R  R  P  A  P  G  G  A  A  A  E  G -

         ggggcaggggcaaagagaggccggggagcaaggaggaggtggatgaagaccgcgacgtc
     301 ---------+---------+---------+---------+---------+---------+ 360
         ccccgtcccccgtttctctccggcccctcgttcctcctccacctacttctggcgctgcag

a        G  A  G  G  K  E  R  P  G  S  K  E  E  V  D  E  D  R  D  V  -
b         G  Q  G  A  K  R  G  R  G  A  R  R  R  W  M  K  T  A  T  S  -
c          G  R  G  Q  R  E  A  G  E  Q  G  G  G  G  *  R  P  R  R  R -

         gatgagtcctcccccaagactcccctccctccaaggcctccccagcccaagatgggcgg
     361 ---------+---------+---------+---------+---------+---------+ 420
         ctactcaggaggggggttctgaggggagggaggttccggaggggtcgggttctacccgcc

a        D  E  S  S  P  Q  D  S  P  P  S  K  A  S  P  A  Q  D  G  R  -
b         M  S  P  P  P  K  T  P  L  P  P  R  P  P  Q  P  K  M  G  G  -
c          *  V  L  P  P  R  L  P  S  L  Q  G  L  P  S  P  R  W  A  A -

         cctccccagacagccgccagagaagccaccagcatcccaggcttcccagcggagggtgcc
     421 ---------+---------+---------+---------+---------+---------+ 480
         ggaggggtctgtcggcggtctcttcggtggtcgtagggtccgaagggtcgcctcccacgg

a        P  P  Q  T  A  A  R  E  A  T  S  I  P  G  F  P  A  E  G  A  -
b         L  P  R  Q  P  P  E  K  P  P  A  S  Q  A  S  Q  R  R  V  P  -
c          S  P  D  S  R  Q  R  S  H  Q  H  P  R  L  P  S  G  G  C  H -

         atccccctccctgtggatttcctctccaaagtttccacagagatcccagcctcagagccc
     481 ---------+---------+---------+---------+---------+---------+ 540
         taggggagggacacctaaaggagaggtttcaaaggtgtctctagggtcggagtctcggg

a        I  P  L  P  V  D  F  L  S  K  V  S  T  E  I  P  A  S  E  P  -
b         S  P  S  L  W  I  S  S  P  K  F  P  Q  R  S  Q  P  Q  S  P  -
c          P  P  P  C  G  F  P  L  Q  S  F  H  R  D  P  S  L  R  A  R -
```

Figure 12-3

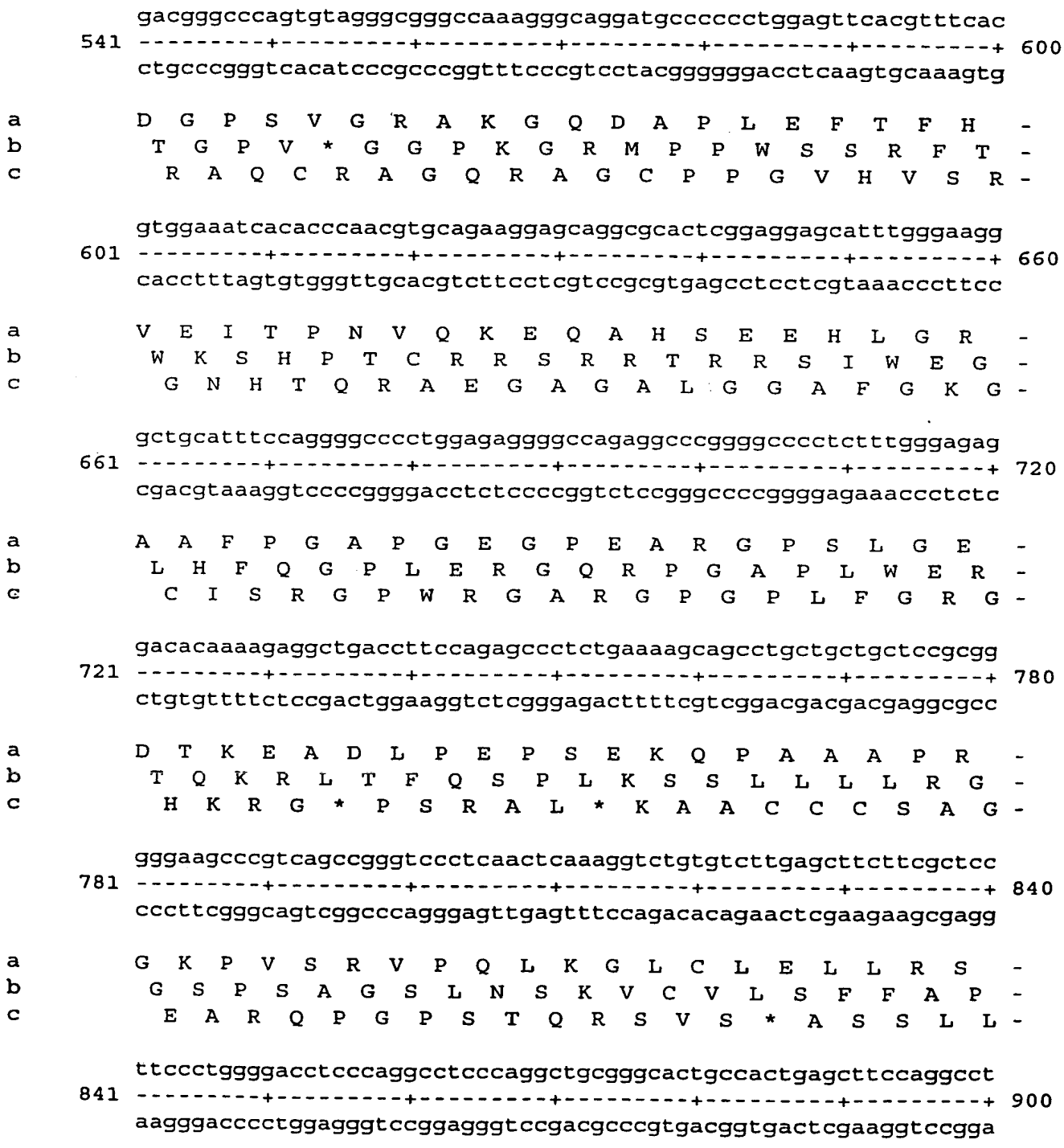
gacgggcccagtgtagggcgggccaaagggcaggatgccccccctggagttcacgtttcac
541 ---------+---------+---------+---------+---------+---------+ 600
ctgcccgggtcacatcccgcccggtttcccgtcctacgggggggacctcaagtgcaaagtg
a D G P S V G R A K G Q D A P L E F T F H -
b T G P V * G G P K G R M P P W S S R F T -
c R A Q C R A G Q R A G C P P G V H V S R -
gtggaaatcacacccaacgtgcagaaggagcaggcgcactcggaggagcatttgggaagg
601 ---------+---------+---------+---------+---------+---------+ 660
caccttagtgtgggttgcacgtcttcctcgtccgcgtgagcctcctcgtaaacccttcc
a V E I T P N V Q K E Q A H S E E H L G R -
b W K S H P T C R R S R R T R R S I W E G -
c G N H T Q R A E G A G A L G G A F G K G -
gctgcatttccaggggcccctggagaggggccagaggcccggggcccctctttgggagag
661 ---------+---------+---------+---------+---------+---------+ 720
cgacgtaaaggtccccggggacctctccccggtctccgggccccggggagaaaccctctc
a A A F P G A P G E G P E A R G P S L G E -
b L H F Q G P L E R G Q R P G A P L W E R -
c C I S R G P W R G A R G P G P L F G R G -
gacacaaaagaggctgaccttccagagccctctgaaaagcagcctgctgctgctccgcgg
721 ---------+---------+---------+---------+---------+---------+ 780
ctgtgttttctccgactggaaggtctcgggagacttttcgtcggacgacgacgaggcgcc
a D T K E A D L P E P S E K Q P A A A P R -
b T Q K R L T F Q S P L K S S L L L L R G -
c H K R G * P S R A L * K A A C C C S A G -
gggaagcccgtcagccgggtccctcaactcaaaggtctgtgtcttgagcttcttcgctcc
781 ---------+---------+---------+---------+---------+---------+ 840
cccttcgggcagtcggcccagggagttgagtttccagacacagaactcgaagaagcgagg
a G K P V S R V P Q L K G L C L E L L R S -
b G S P S A G S L N S K V C V L S F F A P -
c E A R Q P G P S T Q R S V S * A S S L L -
ttccctggggacctcccaggcctcccaggctgcgggcactgccactgagcttccaggcct
841 ---------+---------+---------+---------+---------+---------+ 900
aagggacccctggagggtccggagggtccgacgcccgtgacggtgactcgaaggtccgga

Figure 12-4

```
a        F  P  G  D  L  P  G  L  P  G  C  G  H  C  H  *  A  S  R  P  -
b         S  L  G  T  S  Q  A  S  Q  A  A  G  T  A  T  E  L  P  G  L  -
c          P  W  G  P  P  R  P  P  R  L  R  A  L  P  L  S  F  Q  A  S  -

         cccgactcctgctgcttctgacgttcctaggacgccactaaatcgacacctggg
     901 ---------+---------+---------+---------+---------+---- 954
         gggctgaggacgacgaagactgcaaggatcctgcggtgatttagctgtggaccc

a        P  D  S  C  C  F  *  R  S  *  D  A  T  K  S  T  P  G   -
b         P  T  P  A  A  S  D  V  P  R  T  P  L  N  R  H  L     -
c          R  L  L  L  L  L  T  F  L  G  R  H  *  I  D  T  W    -
```

Enzymes that do cut:

NotI

Enzymes that do not cut:

NONE

Figure 13-1 GFAP

(Linear) MAP of: hsgfap check: 1566 from: 1 to: 3017

RL;HSGFAP - Human glial fibrillary acidic protein (GFAP) mRNA, complete cds.
ID HSGFAP standard; RNA; HUM; 3017 BP.
AC J04569;
NI g183074
DT 23-APR-1990 (Rel. 23, Created)
DT 16-DEC-1994 (Rel. 42, Last updated, Version 3) . . .

With 1 enzymes: NOTI

```
          ccgatggagaggagacgcatcacctccgctgctcgccgctcctacgtctcctcaggggag
        1 ---------+---------+---------+---------+---------+---------+ 60
          ggctacctctcctctgcgtagtggaggcgacgagcggcgaggatgcagaggagtcccctc

a         P  M  E  R  R  R  I  T  S  A  A  R  R  S  Y  V  S  S  G  E  -
b          R  W  R  G  D  A  S  P  P  L  L  A  A  P  T  S  P  Q  G  R  -
c           D  G  E  E  T  H  H  L  R  C  S  P  L  L  R  L  L  R  G  D -

          atgatggtgggggcctggctcctggccgccgtctgggtcctggcacccgcctctccctg
       61 ---------+---------+---------+---------+---------+---------+ 120
          tactaccacccccggaccgaggaccggcggcagacccaggaccgtgggcggagagggac

a         M  M  V  G  G  L  A  P  G  R  R  L  G  P  G  T  R  L  S  L  -
b          *  W  W  G  A  W  L  L  A  A  V  W  V  L  A  P  A  S  P  W  -
c           D  G  G  G  P  G  S  W  P  P  S  G  S  W  H  P  P  L  P  G -

          gctcgaatgccccctccactcccgacccgggtggatttctccctggctggggcactcaat
      121 ---------+---------+---------+---------+---------+---------+ 180
          cgagcttacgggggaggtgagggctgggcccacctaaagagggaccgaccccgtgagtta

a         A  R  M  P  P  P  L  P  T  R  V  D  F  S  L  A  G  A  L  N  -
b          L  E  C  P  L  H  S  R  P  G  W  I  S  P  W  L  G  H  S  M  -
c           S  N  A  P  S  T  P  D  P  G  G  F  L  P  G  W  G  T  Q  C -

          gctggcttcaaggagacccgggccagtgagcgggcagagatgatggagctcaatgaccgc
      181 ---------+---------+---------+---------+---------+---------+ 240
          cgaccgaagttcctctgggcccggtcactcgcccgtctctactacctcgagttactggcg

a         A  G  F  K  E  T  R  A  S  E  R  A  E  M  M  E  L  N  D  R  -
b          L  A  S  R  R  P  G  P  V  S  G  Q  R  *  W  S  S  M  T  A  -
c           W  L  Q  G  D  P  G  Q  *  A  G  R  D  D  G  A  Q  *  P  L -
```

Figure 13-2

```
      tttgccagctacatcgagaaggttcgcttcctggaacagcaaaacaaggcgctggctgct
  241 ---------+---------+---------+---------+---------+---------+ 300
      aaacggtcgatgtagctcttccaagcgaaggaccttgtcgttttgttccgcgaccgacga

a         F  A  S  Y  I  E  K  V  R  F  L  E  Q  Q  N  K  A  L  A  A  -
b          L  P  A  T  S  R  R  F  A  S  W  N  S  K  T  R  R  W  L  L  -
c           C  Q  L  H  R  E  G  S  L  P  G  T  A  K  Q  G  A  G  C  *  -

      gagctgaaccagctgcgggccaaggagcccaccaagctggcagacgtctaccaggctgag
  301 ---------+---------+---------+---------+---------+---------+ 360
      ctcgacttggtcgacgcccggttcctcgggtggttcgaccgtctgcagatggtccgactc

a         E  L  N  Q  L  R  A  K  E  P  T  K  L  A  D  V  Y  Q  A  E  -
b          S  *  T  S  C  G  P  R  S  P  P  S  W  Q  T  S  T  R  L  S  -
c           A  E  P  A  A  G  Q  G  A  H  Q  A  G  R  R  L  P  G  *  A  -

      ctgcgagagctgcggctgcggctcgatcaactcaccgccaacagcgcccggctggaggtt
  361 ---------+---------+---------+---------+---------+---------+ 420
      gacgctctcgacgccgacgccgagctagttgagtggcggttgtcgcgggccgacctccaa

a         L  R  E  L  R  L  R  L  D  Q  L  T  A  N  S  A  R  L  E  V  -
b          C  E  S  C  G  C  G  S  I  N  S  P  P  T  A  P  G  W  R  L  -
c           A  R  A  A  A  A  A  R  S  T  H  R  Q  Q  R  P  A  G  G  *  -

      gagagggacaatctggcacaggacctggccactgtgaggcagaagctccaggatgaaacc
  421 ---------+---------+---------+---------+---------+---------+ 480
      ctctccctgttagaccgtgtcctggaccggtgacactccgtcttcgaggtcctactttgg

a         E  R  D  N  L  A  Q  D  L  A  T  V  R  Q  K  L  Q  D  E  T  -
b          R  G  T  I  W  H  R  T  W  P  L  *  G  R  S  S  R  M  K  P  -
c           E  G  Q  S  G  T  G  P  G  H  C  E  A  E  A  P  G  *  N  Q  -

      aacctgaggctggaagccgagaacaacctggctgcctatagacaggaagcagatgaagcc
  481 ---------+---------+---------+---------+---------+---------+ 540
      ttggactccgaccttcggctcttgttggaccgacggatatctgtccttcgtctacttcgg

a         N  L  R  L  E  A  E  N  N  L  A  A  Y  R  Q  E  A  D  E  A  -
b          T  *  G  W  K  P  R  T  T  W  L  P  I  D  R  K  Q  M  K  P  -
c           P  E  A  G  S  R  E  Q  P  G  C  L  *  T  G  S  R  *  S  H  -

      accctggcccgtctggatctggagaggaagattgagtcgctggaggaggagatccggttc
  541 ---------+---------+---------+---------+---------+---------+ 600
      tgggaccgggcagacctagacctctccttctaactcagcgacctcctcctctaggccaag
```

Figure 13-3

```
a        T  L  A  R  L  D  L  E  R  K  I  E  S  L  E  E  E  I  R  F  -
b         P  W  P  V  W  I  W  R  G  R  L  S  R  W  R  R  R  S  G  S  -
c          P  G  P  S  G  S  G  E  E  D  *  V  A  G  G  G  D  P  V  L -

         ttgaggaagatccacgaggaggaggttcgggaactccaggagcagctggcccgacagcag
     601 ---------+---------+---------+---------+---------+---------+ 660
         aactccttctaggtgctcctcctccaagcccttgaggtcctcgtcgaccgggctgtcgtc

a        L  R  K  I  H  E  E  E  V  R  E  L  Q  E  Q  L  A  R  Q  Q  -
b         *  G  R  S  T  R  R  R  F  G  N  S  R  S  S  W  P  D  S  R  -
c          E  E  D  P  R  G  G  G  S  G  T  P  G  A  A  G  P  T  A  G -

         gtccatgtggagcttgacgtggccaagccagacctcaccgcagccctgaaagagatccgc
     661 ---------+---------+---------+---------+---------+---------+ 720
         caggtacacctcgaactgcaccggttcggtctggagtggcgtcgggactttctctaggcg

a        V  H  V  E  L  D  V  A  K  P  D  L  T  A  A  L  K  E  I  R  -
b         S  M  W  S  L  T  W  P  S  Q  T  S  P  Q  P  *  K  R  S  A  -
c          P  C  G  A  *  R  G  Q  A  R  P  H  R  S  P  E  R  D  P  H -

         acgcagtatgaggcaatggcgtccagcaacatgcatgaagccgaagagtggtaccgctcc
     721 ---------+---------+---------+---------+---------+---------+ 780
         tgcgtcatactccgttaccgcaggtcgttgtacgtacttcggcttctcaccatggcgagg

a        T  Q  Y  E  A  M  A  S  S  N  M  H  E  A  E  E  W  Y  R  S  -
b         R  S  M  R  Q  W  R  P  A  T  C  M  K  P  K  S  G  T  A  P  -
c          A  V  *  G  N  G  V  Q  Q  H  A  *  S  R  R  V  V  P  L  Q -

         aagtttgcagacctgacagacgctgctgcccgcaacgcggagctgctccgccaggccaag
     781 ---------+---------+---------+---------+---------+---------+ 840
         ttcaaacgtctggactgtctgcgacgacgggcgttgcgcctcgacgaggcggtccggttc

a        K  F  A  D  L  T  D  A  A  A  R  N  A  E  L  L  R  Q  A  K  -
b         S  L  Q  T  *  Q  T  L  L  P  A  T  R  S  C  S  A  R  P  S  -
c          V  C  R  P  D  R  R  C  C  P  Q  R  G  A  A  P  P  G  Q  A -

         cacgaagccaacgactaccggcgccagttgcagtccttgacctgcgacctggagtctctg
     841 ---------+---------+---------+---------+---------+---------+ 900
         gtgcttcggttgctgatggccgcggtcaacgtcaggaactggacgctggacctcagagac

a        H  E  A  N  D  Y  R  R  Q  L  Q  S  L  T  C  D  L  E  S  L  -
b         T  K  P  T  T  T  G  A  S  C  S  P  *  P  A  T  W  S  L  C  -
c          R  S  Q  R  L  P  A  P  V  A  V  L  D  L  R  P  G  V  S  A -
```

Figure 13-4

```
        cgcggcacgaacgagtccctggagaggcagatgcgcgagcaggaggagcggcacgtgcgg
    901 ---------+---------+---------+---------+---------+---------+ 960
        gcgccgtgcttgctcagggacctctccgtctacgcgctcgtcctcctcgccgtgcacgcc

a       R  G  T  N  E  S  L  E  R  Q  M  R  E  Q  E  E  R  H  V  R  -
b        A  A  R  T  S  P  W  R  G  R  C  A  S  R  R  S  G  T  C  G -
c         R  H  E  R  V  P  G  E  A  D  A  R  A  G  G  A  A  R  A  G -

        gaggcggccagttatcaggaggcgctggcgcggctggaggaagaggggcagagcctcaag
    961 ---------+---------+---------+---------+---------+---------+ 1020
        ctccgccggtcaatagtcctccgcgaccgcgccgacctccttctccccgtctcggagttc

a       E  A  A  S  Y  Q  E  A  L  A  R  L  E  E  E  G  Q  S  L  K  -
b        R  R  P  V  I  R  R  R  W  R  G  W  R  K  R  G  R  A  S  R -
c         G  G  Q  L  S  G  G  A  G  A  A  G  G  R  G  A  E  P  Q  G -

        gacgagatggcccgccacttgcaggagtaccaggacctgctcaatgtcaagctggccctg
   1021 ---------+---------+---------+---------+---------+---------+ 1080
        ctgctctaccgggcggtgaacgtcctcatggtcctggacgagttacagttcgaccgggac

a       D  E  M  A  R  H  L  Q  E  Y  Q  D  L  L  N  V  K  L  A  L  -
b        T  R  W  P  A  T  C  R  S  T  R  T  C  S  M  S  S  W  P  W -
c         R  D  G  P  P  L  A  G  V  P  G  P  A  Q  C  Q  A  G  P  G -

        gacatcgagatcgccacctacaggaagctgctagagggcgaggagaaccggatcaccatt
   1081 ---------+---------+---------+---------+---------+---------+ 1140
        ctgtagctctagcggtggatgtccttcgacgatctcccgctcctcttggcctagtggtaa

a       D  I  E  I  A  T  Y  R  K  L  L  E  G  E  E  N  R  I  T  I  -
b        T  S  R  S  P  P  T  G  S  C  *  R  A  R  R  T  G  S  P  F -
c         H  R  D  R  H  L  Q  E  A  A  R  G  R  G  E  P  D  H  H  S -

        cccgtgcagaccttctccaacctgcagattcgagaaaccagcctggacaccaagtctgtg
   1141 ---------+---------+---------+---------+---------+---------+ 1200
        gggcacgtctggaagaggttggacgtctaagctctttggtcggacctgtggttcagacac

a       P  V  Q  T  F  S  N  L  Q  I  R  E  T  S  L  D  T  K  S  V  -
b        P  C  R  P  S  P  T  C  R  F  E  K  P  A  W  T  P  S  L  C -
c         R  A  D  L  L  Q  P  A  D  S  R  N  Q  P  G  H  Q  V  C  V -

        tcagaaggccacctcaagaggaacatcgtggtgaagaccgtggagatgcgggatggagag
   1201 ---------+---------+---------+---------+---------+---------+ 1260
        agtcttccggtggagttctccttgtagcaccacttctggcacctctacgccctacctctc
```

Figure 13-5

```
a        S  E  G  H  L  K  R  N  I  V  V  K  T  V  E  M  R  D  G  E  -
b         Q  K  A  T  S  R  G  T  S  W  *  R  P  W  R  C  G  M  E  R -
c          R  R  P  P  Q  E  E  H  R  G  E  D  R  G  D  A  G  W  R  G-

         gtcattaaggagtccaagcaggagcacaaggatgtgatgtgaggcaggacccacctggtg
    1261 ---------+---------+---------+---------+---------+---------+ 1320
         cagtaattcctcaggttcgtcctcgtgttcctacactacactccgtcctgggtggaccac

a        V  I  K  E  S  K  Q  E  H  K  D  V  M  *  G  R  T  H  L  V  -
b         S  L  R  S  P  S  R  S  T  R  M  *  C  E  A  G  P  T  W  W -
c          H  *  G  V  Q  A  G  A  Q  G  C  D  V  R  Q  D  P  P  G  G-

         gcctctgccccgtctcatgaggggcccgagcagaagcaggatagttgctccgcctctgct
    1321 ---------+---------+---------+---------+---------+---------+ 1380
         cggagacggggcagagtactccccgggctcgtcttcgtcctatcaacgaggcggagacga

a        A  S  A  P  S  H  E  G  P  E  Q  K  Q  D  S  C  S  A  S  A  -
b         P  L  P  R  L  M  R  G  P  S  R  S  R  I  V  A  P  P  L  L -
c          L  C  P  V  S  *  G  A  R  A  E  A  G  *  L  L  R  L  C  W-

         ggcacatttccccagacctgagctccccaccaccccagctgctcccctccctcctctgtc
    1381 ---------+---------+---------+---------+---------+---------+ 1440
         ccgtgtaaaggggtctggactcgaggggtggtggggtcgacgaggggagggaggagacag

a        G  T  F  P  Q  T  *  A  P  H  H  P  S  C  S  P  P  S  S  V  -
b         A  H  F  P  R  P  E  L  P  T  T  P  A  A  P  L  P  P  L  S -
c          H  I  S  P  D  L  S  S  P  P  P  Q  L  L  P  S  L  L  C  P-

         cctaggtcagcttgctgccctaggctccgtcagtatcaggcctgccagacggcacccacc
    1441 ---------+---------+---------+---------+---------+---------+ 1500
         ggatccagtcgaacgacgggatccgaggcagtcatagtccggacggtctgccgtgggtgg

a        P  R  S  A  C  C  P  R  L  R  Q  Y  Q  A  C  Q  T  A  P  T  -
b         L  G  Q  L  A  A  L  G  S  V  S  I  R  P  A  R  R  H  P  P -
c          *  V  S  L  L  P  *  A  P  S  V  S  G  L  P  D  G  T  H  P-

         cagcacccagcaactccaactaacaagaaactcacccccaagggcagtctggaggggcat
    1501 ---------+---------+---------+---------+---------+---------+ 1560
         gtcgtgggtcgttgaggttgattgttctttgagtgggggttcccgtcagacctccccgta

a        Q  H  P  A  T  P  T  N  K  K  L  T  P  K  G  S  L  E  G  H  -
b         S  T  Q  Q  L  Q  L  T  R  N  S  P  P  R  A  V  W  R  G  M -
c          A  P  S  N  S  N  *  Q  E  T  H  P  Q  G  Q  S  G  G  A  W-
```

Figure 13-6

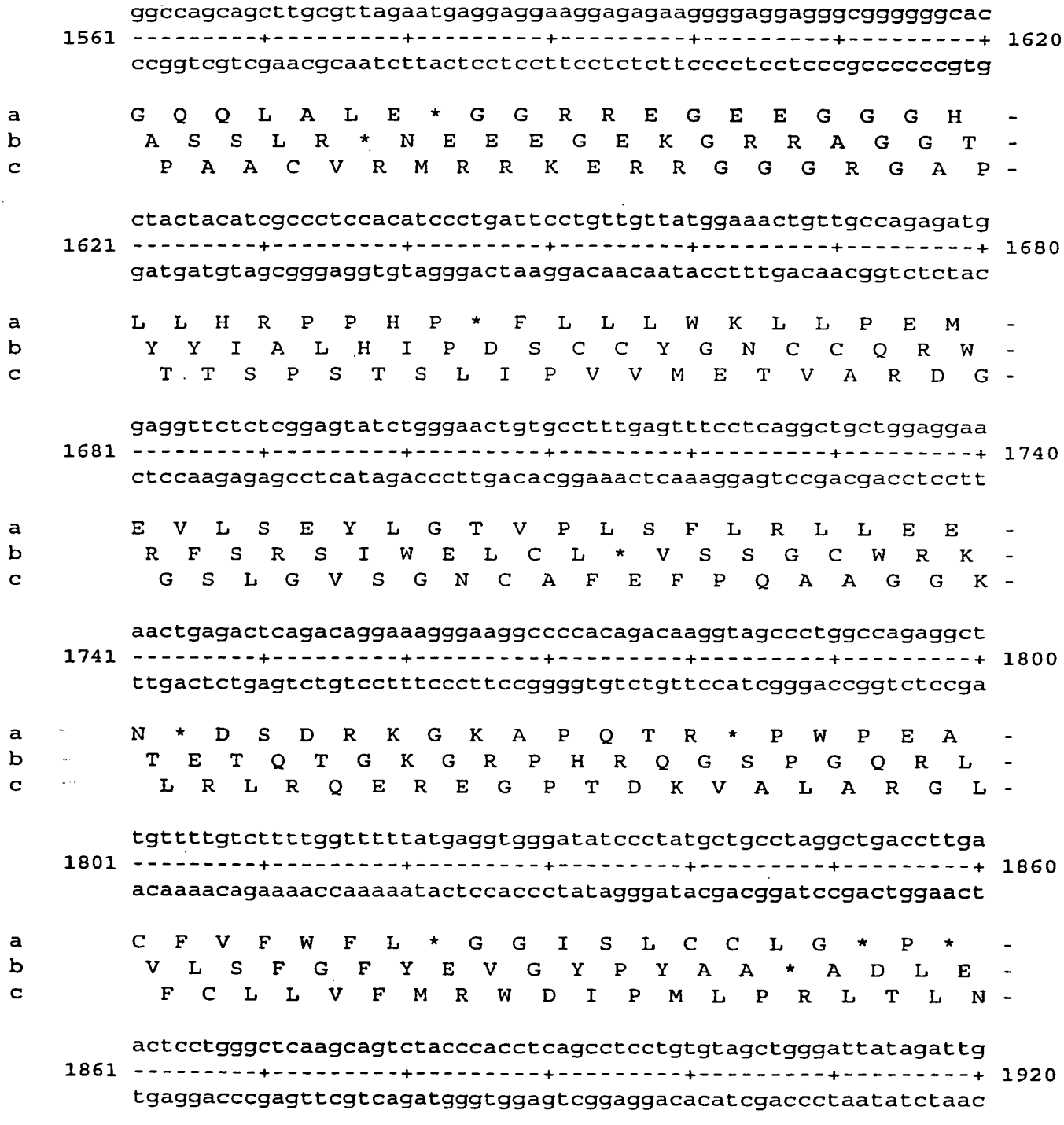
ggccagcagcttgcgttagaatgaggaggaaggagagaaggggaggagggcgggggcac
1561 ---------+---------+---------+---------+---------+---------+---------+ 1620
ccggtcgtcgaacgcaatcttactcctccttcctctcttcccctcctcccgcccccgtg
a G Q Q L A L E * G G R R E G E E G G G H -
b A S S L R * N E E E G E K G R R A G G T -
c P A A C V R M R R K E R R G G G R G A P -
ctactacatcgccctccacatccctgattcctgttgttatggaaactgttgccagagatg
1621 ---------+---------+---------+---------+---------+---------+---------+ 1680
gatgatgtagcgggaggtgtagggactaaggacaacaataccttgacaacggtctctac
a L L H R P P H P * F L L L W K L L P E M -
b Y Y I A L H I P D S C C Y G N C C Q R W -
c T T S P S T S L I P V V M E T V A R D G -
gaggttctctcggagtatctgggaactgtgcctttgagtttcctcaggctgctggaggaa
1681 ---------+---------+---------+---------+---------+---------+---------+ 1740
ctccaagagagcctcatagacccttgacacggaaactcaaaggagtccgacgacctcctt
a E V L S E Y L G T V P L S F L R L L E E -
b R F S R S I W E L C L * V S S G C W R K -
c G S L G V S G N C A F E F P Q A A G G K -
aactgagactcagacaggaaagggaaggccccacagacaaggtagccctggccagaggct
1741 ---------+---------+---------+---------+---------+---------+---------+ 1800
ttgactctgagtctgtcctttcccttccggggtgtctgttccatcgggaccggtctccga
a N * D S D R K G K A P Q T R * P W P E A -
b T E T Q T G K G R P H R Q G S P G Q R L -
c L R L R Q E R E G P T D K V A L A R G L -
tgttttgtcttttggtttttatgaggtgggatatccctatgctgcctaggctgaccttga
1801 ---------+---------+---------+---------+---------+---------+---------+ 1860
acaaaacagaaaaccaaaaatactccaccctatagggatacgacggatccgactggaact
a C F V F W F L * G G I S L C C L G * P * -
b V L S F G F Y E V G Y P Y A A * A D L E -
c F C L L V F M R W D I P M L P R L T L N -
actcctgggctcaagcagtctacccacctcagcctcctgtgtagctgggattatagattg
1861 ---------+---------+---------+---------+---------+---------+---------+ 1920
tgaggacccgagttcgtcagatgggtggagtcggaggacacatcgaccctaatatctaac

Figure 13-7

```
a        T  P  G  L  K  Q  S  T  H  L  S  L  L  C  S  W  D  Y  R  L  -
b         L  L  G  S  S  S  L  P  T  S  A  S  C  V  A  G  I  I  D  W  -
c          S  W  A  Q  A  V  Y  P  P  Q  P  P  V  *  L  G  L  *  I  G -

         gagccaccatgcccagctcagagggttgttctcctagactgaccctgatcagtctaagat
    1921 ---------+---------+---------+---------+---------+---------+ 1980
         ctcggtggtacgggtcgagtctcccaacaagaggatctgactgggactagtcagattcta

a        E  P  P  C  P  A  Q  R  V  V  L  L  D  *  P  *  S  V  *  D  -
b         S  H  H  A  Q  L  R  G  L  F  S  *  T  D  P  D  Q  S  K  M  -
c          A  T  M  P  S  S  E  G  C  S  P  R  L  T  L  I  S  L  R  W -

         gggtggggacgtcctgccacctggggcagtcacctgcccagatcccagaaggacctcctg
    1981 ---------+---------+---------+---------+---------+---------+ 2040
         cccacccctgcaggacggtggaccccgtcagtggacgggtctagggtcttcctggaggac

a        G  W  G  R  P  A  T  W  G  S  H  L  P  R  S  Q  K  D  L  L  -
b         G  G  D  V  L  P  P  G  A  V  T  C  P  D  P  R  R  T  S  *  -
c          V  G  T  S  C  H  L  G  Q  S  P  A  Q  I  P  E  G  P  P  E -

         agcgatgactcaagtgtctcagtccacctgagctgccatccagggatgccatctgtgggc
    2041 ---------+---------+---------+---------+---------+---------+ 2100
         tcgctactgagttcacagagtcaggtggactcgacggtaggtccctacggtagacacccg

a        S  D  D  S  S  V  S  V  H  L  S  C  H  P  G  M  P  S  V  G  -
b         A  M  T  Q  V  S  Q  S  T  *  A  A  I  Q  G  C  H  L  W  A  -
c          R  *  L  K  C  L  S  P  P  E  L  P  S  R  D  A  I  C  G  H -

         acgctgtgggcaggtgggagcttgattctcagcacttgggggatctgttgtgtacgtgga
    2101 ---------+---------+---------+---------+---------+---------+ 2160
         tgcgacacccgtccaccctcgaactaagagtcgtgaacccctagacaacacatgcacct

a        T  L  W  A  G  G  S  L  I  L  S  T  W  G  I  C  C  V  R  G  -
b         R  C  G  Q  V  G  A  *  F  S  A  L  G  G  S  V  V  Y  V  E  -
c          A  V  G  R  W  E  L  D  S  Q  H  L  G  D  L  L  C  T  W  R -

         gagggatgaggtgctgggagggatagaggggggctgcctggcccccagctgtgggtacag
    2161 ---------+---------+---------+---------+---------+---------+ 2220
         ctccctactccacgaccctccctatctcccccgacggaccgggggtcgacacccatgtc

a        E  G  *  G  A  G  R  D  R  G  G  L  P  G  P  Q  L  W  V  Q  -
b         R  D  E  V  L  G  G  I  E  G  G  C  L  A  P  S  C  G  Y  R  -
c          G  M  R  C  W  E  G  *  R  G  A  A  W  P  P  A  V  G  T  E -
```

Figure 13-8

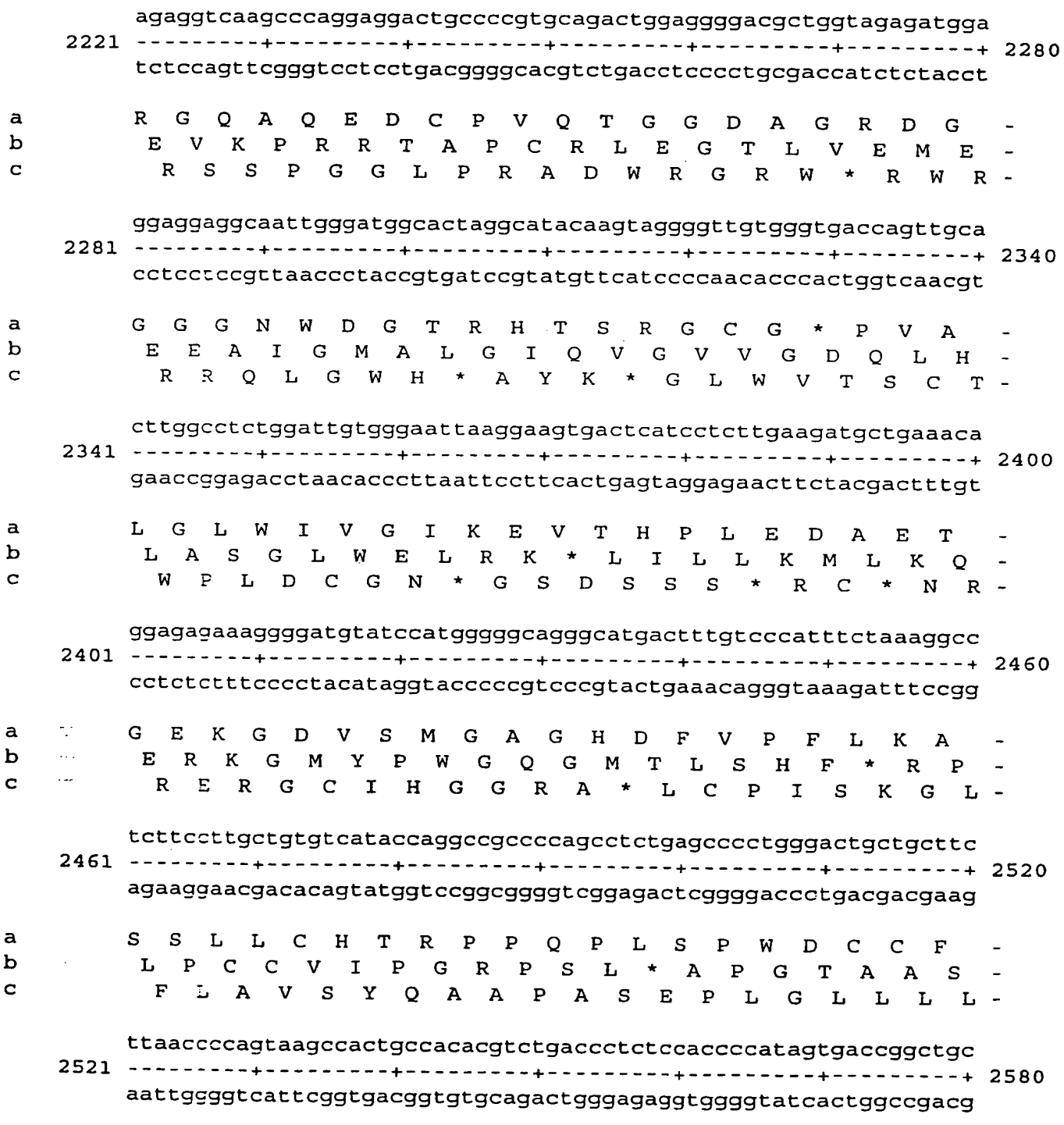
agaggtcaagcccaggaggactgccccgtgcagactggaggggacgctggtagagatgga
2221 ---------+---------+---------+---------+---------+---------+---------+ 2280
tctccagttcgggtcctcctgacggggcacgtctgacctcccctgcgaccatctctacct
a R G Q A Q E D C P V Q T G G D A G R D G -
b E V K P R R T A P C R L E G T L V E M E -
c R S S P G G L P R A D W R G R W * R W R -
ggaggaggcaattgggatggcactaggcatacaagtaggggttgtgggtgaccagttgca
2281 ---------+---------+---------+---------+---------+---------+---------+ 2340
cctcctccgttaaccctaccgtgatccgtatgttcatccccaacacccactggtcaacgt
a G G G N W D G T R H T S R G C G * P V A -
b E E A I G M A L G I Q V G V V G D Q L H -
c R R Q L G W H * A Y K * G L W V T S C T -
cttggcctctggattgtgggaattaaggaagtgactcatcctcttgaagatgctgaaaca
2341 ---------+---------+---------+---------+---------+---------+---------+ 2400
gaaccggagacctaacacccttaattccttcactgagtaggagaacttctacgactttgt
a L G L W I V G I K E V T H P L E D A E T -
b L A S G L W E L R K * L I L L K M L K Q -
c W P L D C G N * G S D S S S * R C * N R -
ggagagaaaggggatgtatccatgggggcagggcatgactttgtcccatttctaaaggcc
2401 ---------+---------+---------+---------+---------+---------+---------+ 2460
cctctctttcccctacataggtacccccgtcccgtactgaaacagggtaaagatttccgg
a G E K G D V S M G A G H D F V P F L K A -
b E R K G M Y P W G Q G M T L S H F * R P -
c R E R G C I H G G R A * L C P I S K G L -
tcttccttgctgtgtcataccaggccgccccagcctctgagcccctgggactgctgcttc
2461 ---------+---------+---------+---------+---------+---------+---------+ 2520
agaaggaacgacacagtatggtccggcggggtcggagactcggggaccctgacgacgaag
a S S L L C H T R P P Q P L S P W D C C F -
b L P C C V I P G R P S L * A P G T A A S -
c F L A V S Y Q A A P A S E P L G L L L L -
ttaaccccagtaagccactgccacacgtctgaccctctccaccccatagtgaccggctgc
2521 ---------+---------+---------+---------+---------+---------+---------+ 2580
aattggggtcattcggtgacggtgtgcagactgggagaggtggggtatcactggccgacg

Figure 13-9

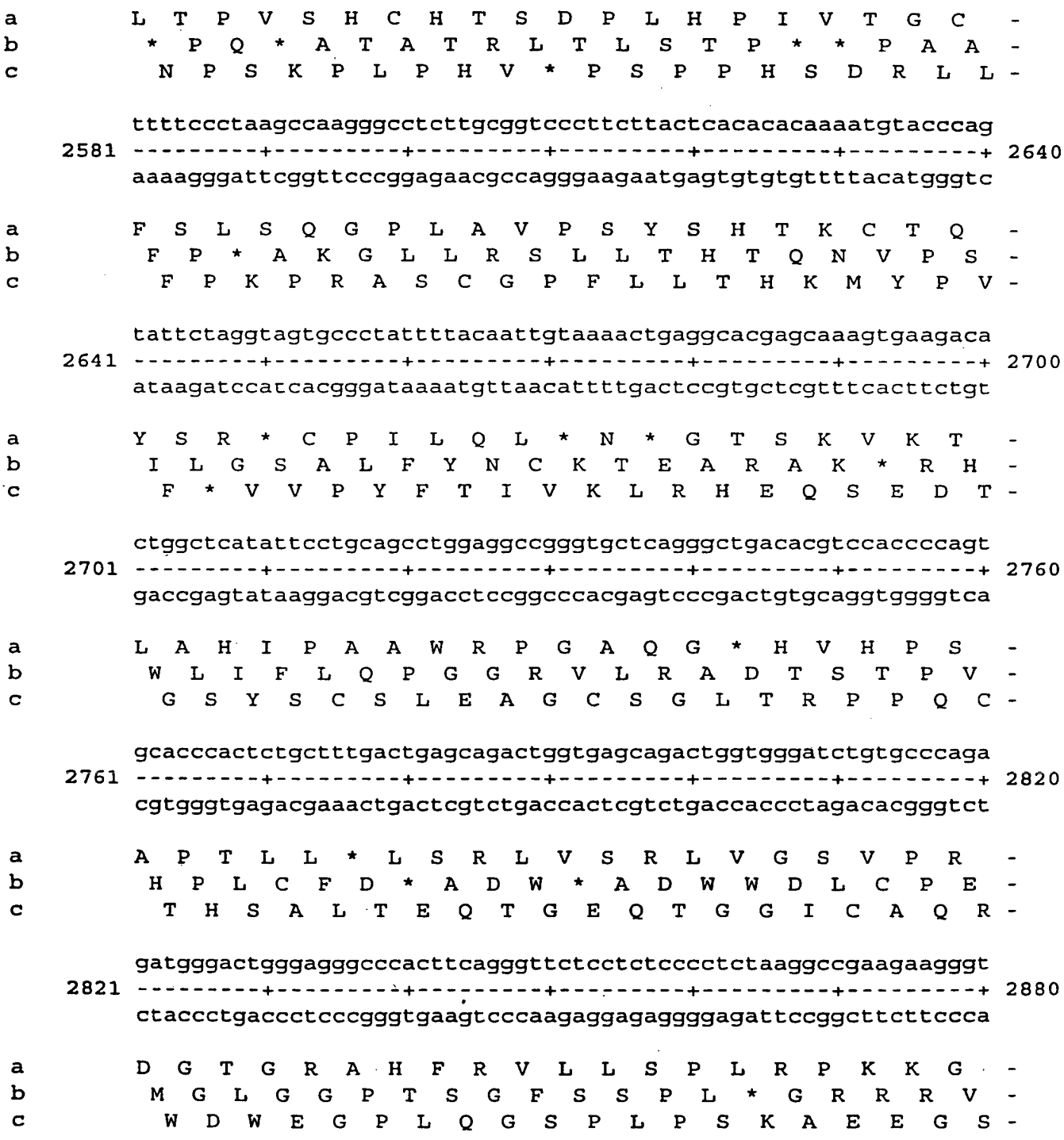
a L T P V S H C H T S D P L H P I V T G C -
b * P Q * A T A T R L T L S T P * * P A A -
c N P S K P L P H V * P S P P H S D R L L -
ttttccctaagccaagggcctcttgcggtcccttcttactcacacacaaaatgtacccag
2581 2640
aaaagggattcggttcccggagaacgccagggaagaatgagtgtgtgttttacatgggtc
a F S L S Q G P L A V P S Y S H T K C T Q -
b F P * A K G L L R S L L T H T Q N V P S -
c F P K P R A S C G P F L L T H K M Y P V -
tattctaggtagtgccctattttacaattgtaaaactgaggcacgagcaaagtgaagaca
2641 2700
ataagatccatcacgggataaaatgttaacattttgactccgtgctcgtttcacttctgt
a Y S R * C P I L Q L * N * G T S K V K T -
b I L G S A L F Y N C K T E A R A K * R H -
c F * V V P Y F T I V K L R H E Q S E D T -
ctggctcatattcctgcagcctggaggccgggtgctcagggctgacacgtccaccccagt
2701 2760
gaccgagtataaggacgtcggacctccggcccacgagtcccgactgtgcaggtggggtca
a L A H I P A A W R P G A Q G * H V H P S -
b W L I F L Q P G G R V L R A D T S T P V -
c G S Y S C S L E A G C S G L T R P P Q C -
gcacccactctgctttgactgagcagactggtgagcagactggtgggatctgtgcccaga
2761 2820
cgtgggtgagacgaaactgactcgtctgaccactcgtctgaccaccctagacacgggtct
a A P T L L * L S R L V S R L V G S V P R -
b H P L C F D * A D W * A D W W D L C P E -
c T H S A L T E Q T G E Q T G G I C A Q R -
gatgggactgggagggcccacttcagggttctcctctcccctctaaggccgaagaagggt
2821 2880
ctaccctgaccctcccgggtgaagtcccaagaggagaggggagattccggcttcttccca
a D G T G R A H F R V L L S P L R P K K G -
b M G L G G P T S G F S S P L * G R R R V -
c W D W E G P L Q G S P L P S K A E E G S -

Figure 13-10

```
          ccttccctctccccaagacttggtgtcctttccctccacttcttcctgccacctgctgct
     2881 ---------+---------+---------+---------+---------+---------+ 2940
          ggaagggagaggggttctgaaccacaggaaagggaggtgaagaaggacggtggacgacga

a         P  S  L  S  P  R  L  G  V  L  S  L  H  F  F  L  P  P  A  A  -
b          L  P  S  P  Q  D  L  V  S  F  P  S  T  S  S  C  H  L  L  L  -
c           F  P  L  P  K  T  W  C  P  F  P  P  L  L  P  A  T  C  C  C -

          gctgctgctgctaatcttcagggcactgctgctgcctttagtcgctgaggaaaaataaag
     2941 ---------+---------+---------+---------+---------+---------+ 3000
          cgacgacgacgattagaagtcccgtgacgacgacggaaatcagcgactccttttttatttc

a         A  A  A  A  N  L  Q  G  T  A  A  A  F  S  R  *  G  K  I  K  -
b          L  L  L  L  I  F  R  A  L  L  L  P  L  V  A  E  E  K  *  R  -
c           C  C  C  *  S  S  G  H  C  C  C  L  *  S  L  R  K  N  K  D -

          acaaatgctgcgcccttt
     3001 ---------+------- 3017
          tgtttacgacgcgggaa

a         T  N  A  A  P     -
b          Q  M  L  R  P     -
c           K  C  C  A  L     -
```

Enzymes that do cut:

NONE

Enzymes that do not cut:

NotI

Figure 14-1 P53

```
 (Linear) MAP of: hsp53t  check: 640  from: 1  to: 1760

RL;HSP53T     - Human p53 cellular tumor antigen mRNA, complete cds.
ID   HSP53T     standard; RNA; HUM; 1760 BP.
AC   K03199;
NI   g189478
DT   18-NOV-1986 (Rel. 10, Created)
DT   18-JAN-1995 (Rel. 42, Last updated, Version 4) . . .

 With 1 enzymes: NOTI

          gtcgaccctttccacccctggaagatggaaataaacctgcgtgtgggtggagtgttagga
        1 ---------+---------+---------+---------+---------+---------+ 60
          cagctgggaaaggtggggaccttctacctttatttggacgcacacccacctcacaatcct

a         V  D  P  F  H  P  W  K  M  E  I  N  L  R  V  G  G  V  L  G  -
b          S  T  L  S  T  P  G  R  W  K  *  T  C  V  W  V  E  C  *  D -
c           R  P  F  P  P  L  E  D  G  N  K  P  A  C  G  W  S  V  R  T -

          caaaaaaaaaaaaaaaaagtctagagccaccgtccagggagcaggtagctgctgggctc
       61 ---------+---------+---------+---------+---------+---------+ 120
          gttttttttttttttttttcagatctcggtggcaggtccctcgtccatcgacgacccgag

a         Q  K  K  K  K  K  S  L  E  P  P  S  R  E  Q  V  A  A  G  L  -
b          K  K  K  K  K  K  V  *  S  H  R  P  G  S  R  *  L  L  G  S -
c           K  K  K  K  K  K  S  R  A  T  V  Q  G  A  G  S  C  W  A  P -

          cggggacactttgcgttcgggctgggagcgtgctttccacgacggtgacacgcttccctg
      121 ---------+---------+---------+---------+---------+---------+ 180
          gcccctgtgaaacgcaagcccgaccctcgcacgaaaggtgctgccactgtgcgaagggac

a         R  G  H  F  A  F  G  L  G  A  C  F  P  R  R  *  H  A  S  L  -
b          G  D  T  L  R  S  G  W  E  R  A  F  H  D  G  D  T  L  P  W -
c           G  T  L  C  V  R  A  G  S  V  L  S  T  T  V  T  R  F  P  G -

          gattggcagccagactgccttccgggtcactgccatggaggagccgcagtcagatcctag
      181 ---------+---------+---------+---------+---------+---------+ 240
          ctaaccgtcggtctgacggaaggcccagtgacggtacctcctcggcgtcagtctaggatc

a         D  W  Q  P  D  C  L  P  G  H  C  H  G  G  A  A  V  R  S  *  -
b          I  G  S  Q  T  A  F  R  V  T  A  M  E  E  P  Q  S  D  P  S -
c           L  A  A  R  L  P  S  G  S  L  P  W  R  S  R  S  Q  I  L  A -
```

Figure 14-2

```
      cgtcgagccccctctgagtcaggaaacattttcagacctatggaaactacttcctgaaaa
  241 ---------+---------+---------+---------+---------+---------+ 300
      gcagctcgggggagactcagtcctttgtaaaagtctggatacctttgatgaaggactttt

a         R  R  A  P  S  E  S  G  N  I  F  R  P  M  E  T  T  S  *  K  -
b          V  E  P  P  L  S  Q  E  T  F  S  D  L  W  K  L  L  P  E  N -
c           S  S  P  L  *  V  R  K  H  F  Q  T  Y  G  N  Y  F  L  K  T -

      caacgttctgtccccttgccgtcccaagcaatggatgatttgatgctgtccccggacga
  301 ---------+---------+---------+---------+---------+---------+ 360
      gttgcaagacaggggaacggcagggttcgttacctactaaactacgacaggggcctgct

a         Q  R  S  V  P  L  A  V  P  S  N  G  *  F  D  A  V  P  G  R  -
b          N  V  L  S  P  L  P  S  Q  A  M  D  D  L  M  L  S  P  D  D -
c           T  F  C  P  P  C  R  P  K  Q  W  M  I  *  C  C  P  R  T  I -

      tattgaacaatggttcactgaagacccaggtccagatgaagctcccagaatgccagaggc
  361 ---------+---------+---------+---------+---------+---------+ 420
      ataacttgttaccaagtgacttctgggtccaggtctacttcgagggtcttacggtctccg

a         Y  *  T  M  V  H  *  R  P  R  S  R  *  S  S  Q  N  A  R  G  -
b          I  E  Q  W  F  T  E  D  P  G  P  D  E  A  P  R  M  P  E  A -
c           L  N  N  G  S  L  K  T  Q  V  Q  M  K  L  P  E  C  Q  R  L -

      tgctccccccgtggcccctgcaccagcagctcctacaccggcggcccctgcaccagcccc
  421 ---------+---------+---------+---------+---------+---------+ 480
      acgaggggggcaccggggacgtggtcgtcgaggatgtggccgccggggacgtggtcgggg

a         C  S  P  R  G  P  C  T  S  S  S  Y  T  G  G  P  C  T  S  P  -
b          A  P  P  V  A  P  A  P  A  A  P  T  P  A  A  P  A  P  A  P -
c           L  P  P  W  P  L  H  Q  Q  L  L  H  R  R  P  L  H  Q  P  P -

      ctcctggcccctgtcatcttctgtcccttcccagaaaacctaccagggcagctacggttt
  481 ---------+---------+---------+---------+---------+---------+ 540
      gaggaccggggacagtagaagacagggaagggtcttttggatggtcccgtcgatgccaaa

a         L  L  A  P  V  I  F  C  P  F  P  E  N  L  P  G  Q  L  R  F  -
b          S  W  P  L  S  S  S  V  P  S  Q  K  T  Y  Q  G  S  Y  G  F -
c           P  G  P  C  H  L  L  S  L  P  R  K  P  T  R  A  A  T  V  S -

      ccgtctgggcttcttgcattctgggacagccaagtctgtgacttgcacgtactcccctgc
  541 ---------+---------+---------+---------+---------+---------+ 600
      ggcagacccgaagaacgtaagaccctgtcggttcagacactgaacgtgcatgaggggacg
```

Figure 14-3

```
a        P  S  G  L  L  A  F  W  D  S  Q  V  C  D  L  H  V  L  P  C  -
b         R  L  G  F  L  H  S  G  T  A  K  S  V  T  C  T  Y  S  P  A  -
c          V  W  A  S  C  I  L  G  Q  P  S  L  *  L  A  R  T  P  L  P -

         cctcaacaagatgttttgccaactggccaagacctgccctgtgcagctgtgggttgattc
     601 ---------+---------+---------+---------+---------+---------+ 660
         ggagttgttctacaaaacggttgaccggttctggacgggacacgtcgacacccaactaag

a        P  Q  Q  D  V  L  P  T  G  Q  D  L  P  C  A  A  V  G  *  F  -
b         L  N  K  M  F  C  Q  L  A  K  T  C  P  V  Q  L  W  V  D  S  -
c          S  T  R  C  F  A  N  W  P  R  P  A  L  C  S  C  G  L  I  P -

         cacacccccgcccggcacccgcgtccgcgccatggccatctacaagcagtcacagcacat
     661 ---------+---------+---------+---------+---------+---------+ 720
         gtgtgggggcgggccgtgggcgcaggcgcggtaccggtagatgttcgtcagtgtcgtgta

a        H  T  P  A  R  H  P  R  P  R  H  G  H  L  Q  A  V  T  A  H  -
b         T  P  P  P  G  T  R  V  R  A  M  A  I  Y  K  Q  S  Q  H  M  -
c          H  P  R  P  A  P  A  S  A  P  W  P  S  T  S  S  H  S  T  * -

         gacggaggttgtgaggcgctgcccccaccatgagcgctgctcagatagcgatggtctggc
     721 ---------+---------+---------+---------+---------+---------+ 780
         ctgcctccaacactccgcgacgggggtggtactcgcgacgagtctatcgctaccagaccg

a        D  G  G  C  E  A  L  P  P  P  *  A  L  L  R  *  R  W  S  G  -
b         T  E  V  V  R  R  C  P  H  H  E  R  C  S  D  S  D  G  L  A  -
c          R  R  L  *  G  A  A  P  T  M  S  A  A  Q  I  A  M  V  W  P -

         ccctcctcagcatcttatccgagtggaaggaaatttgcgtgtggagtatttggatgacag
     781 ---------+---------+---------+---------+---------+---------+ 840
         gggaggagtcgtagaataggctcaccttcctttaaacgcacacctcataaacctactgtc

a        P  S  S  A  S  Y  P  S  G  R  K  F  A  C  G  V  F  G  *  Q  -
b         P  P  Q  H  L  I  R  V  E  G  N  L  R  V  E  Y  L  D  D  R  -
c          L  L  S  I  L  S  E  W  K  E  I  C  V  W  S  I  W  M  T  E -

         aaacacttttcgacatagtgtggtggtgccctatgagccgcctgaggttggctctgactg
     841 ---------+---------+---------+---------+---------+---------+ 900
         tttgtgaaaagctgtatcacaccaccacgggatactcggcggactccaaccgagactgac

a        K  H  F  S  T  *  C  G  G  A  L  *  A  A  *  G  W  L  *  L  -
b         N  T  F  R  H  S  V  V  V  P  Y  E  P  P  E  V  G  S  D  C  -
c          T  L  F  D  I  V  W  W  C  P  M  S  R  L  R  L  A  L  T  V -
```

Figure 14-4

```
      taccaccatccactacaactacatgtgtaacagttcctgcatgggcggcatgaaccggag
  901 ---------+---------+---------+---------+---------+---------+ 960
      atggtggtaggtgatgttgatgtacacattgtcaaggacgtacccgccgtacttggcctc

a         Y  H  H  P  L  Q  L  H  V  *  Q  F  L  H  G  R  H  E  P  E  -
b          T  T  I  H  Y  N  Y  M  C  N  S  S  C  M  G  G  M  N  R  R  -
c           P  P  S  T  T  T  T  C  V  T  V  P  A  W  A  A  *  T  G  G -

      gcccatcctcaccatcatcacactggaagactccagtggtaatctactgggacggaacag
  961 ---------+---------+---------+---------+---------+---------+ 1020
      cgggtaggagtggtagtagtgtgaccttctgaggtcaccattagatgaccctgccttgtc

a         A  H  P  H  H  H  H  T  G  R  L  Q  W  *  S  T  G  T  E  Q  -
b          P  I  L  T  I  I  T  L  E  D  S  S  G  N  L  L  G  R  N  S  -
c           P  S  S  P  S  S  H  W  K  T  P  V  V  I  Y  W  D  G  T  A -

      ctttgaggtgcatgtttgtgcctgtcctgggagagaccggcgcacagaggaagagaatct
 1021 ---------+---------+---------+---------+---------+---------+ 1080
      gaaactccacgtacaaacacggacaggaccctctctggccgcgtgtctccttctcttaga

a         L  *  G  A  C  L  C  L  S  W  E  R  P  A  H  R  G  R  E  S  -
b          F  E  V  H  V  C  A  C  P  G  R  D  R  R  T  E  E  E  N  L  -
c           L  R  C  M  F  V  P  V  L  G  E  T  G  A  Q  R  K  R  I  S -

      ccgcaagaaaggggagcctcaccacgagctgcccccagggagcactaagcgagcactgcc
 1081 ---------+---------+---------+---------+---------+---------+ 1140
      ggcgttctttcccctcggagtggtgctcgacgggggtccctcgtgattcgctcgtgacgg

a         P  Q  E  R  G  A  S  P  R  A  A  P  R  E  H  *  A  S  T  A  -
b          R  K  K  G  E  P  H  H  E  L  P  P  G  S  T  K  R  A  L  P  -
c           A  R  K  G  S  L  T  T  S  C  P  Q  G  A  L  S  E  H  C  P -

      caacaacaccagctcctctccccagccaaagaagaaaccactggatggagaatatttcac
 1141 ---------+---------+---------+---------+---------+---------+ 1200
      gttgttgtggtcgaggagaggggtcggtttcttctttggtgacctacctcttataaagtg

a         Q  Q  H  Q  L  L  S  P  A  K  E  E  T  T  G  W  R  I  F  H  -
b          N  N  T  S  S  S  P  Q  P  K  K  K  P  L  D  G  E  Y  F  T  -
c           T  T  P  A  P  L  P  S  Q  R  R  N  H  W  M  E  N  I  S  P -

      ccttcagatccgtgggcgtgagcgcttcgagatgttccgagagctgaatgaggccttgga
 1201 ---------+---------+---------+---------+---------+---------+ 1260
      ggaagtctaggcacccgcactcgcgaagctctacaaggctctcgacttactccggaacct
```

Figure 14-5

```
a        P  S  D  P  W  A  *  A  L  R  D  V  P  R  A  E  *  G  L  G  -
b         L  Q  I  R  G  R  E  R  F  E  M  F  R  E  L  N  E  A  L  E  -
c          F  R  S  V  G  V  S  A  S  R  C  S  E  S  *  M  R  P  W  N -

         actcaaggatgcccaggctgggaaggagccaggggggagcagggctcactccagccacct
    1261 ---------+---------+---------+---------+---------+---------+ 1320
         tgagttcctacgggtccgacccttcctcggtccccccctcgtcccgagtgaggtcggtgga

a        T  Q  G  C  P  G  W  E  G  A  R  G  E  Q  G  S  L  Q  P  P  -
b         L  K  D  A  Q  A  G  K  E  P  G  G  S  R  A  H  S  S  H  L  -
c          S  R  M  P  R  L  G  R  S  Q  G  G  A  G  L  T  P  A  T  * -

         gaagtccaaaaagggtcagtctacctcccgccataaaaaactcatgttcaagacagaagg
    1321 ---------+---------+---------+---------+---------+---------+ 1380
         cttcaggtttttcccagtcagatggagggcggtatttttgagtacaagttctgtcttcc

a        E  V  Q  K  G  S  V  Y  L  P  P  *  K  T  H  V  Q  D  R  R  -
b         K  S  K  K  G  Q  S  T  S  R  H  K  K  L  M  F  K  T  E  G  -
c          S  P  K  R  V  S  L  P  P  A  I  K  N  S  C  S  R  Q  K  G -

         gcctgactcagactgacattctccacttcttgttccccactgacagcctcccacccccat
    1381 ---------+---------+---------+---------+---------+---------+ 1440
         cggactgagtctgactgtaagaggtgaagaacaaggggtgactgtcggagggtgggggta

a        A  *  L  R  L  T  F  S  T  S  C  S  P  L  T  A  S  H  P  H  -
b         P  D  S  D  *  H  S  P  L  L  V  P  H  *  Q  P  P  T  P  I  -
c          L  T  Q  T  D  I  L  H  F  L  F  P  T  D  S  L  P  P  P  S -

         ctctccctcccctgccattttgggttttgggtctttgaacccttgcttgcaataggtgtg
    1441 ---------+---------+---------+---------+---------+---------+ 1500
         gagagggaggggacggtaaaacccaaaacccagaaacttgggaacgaacgttatccacac

a        L  S  L  P  C  H  F  G  F  W  V  F  E  P  L  L  A  I  G  V  -
b         S  P  S  P  A  I  L  G  F  G  S  L  N  P  C  L  Q  *  V  C  -
c          L  P  P  L  P  F  W  V  L  G  L  *  T  L  A  C  N  R  C  A -

         cgtcagaagcacccaggacttccatttgctttgtcccggggctccactgaacaagttggc
    1501 ---------+---------+---------+---------+---------+---------+ 1560
         gcagtcttcgtgggtcctgaaggtaaacgaaacagggccccgaggtgacttgttcaaccg

a        R  Q  K  H  P  G  L  P  F  A  L  S  R  G  S  T  E  Q  V  G  -
b         V  R  S  T  Q  D  F  H  L  L  C  P  G  A  P  L  N  K  L  A  -
c          S  E  A  P  R  T  S  I  C  F  V  P  G  L  H  *  T  S  W  P -
```

Figure 14-6

```
      ctgcactggtgttttgttgtggggaggaggatggggagtaggacataccagcttagattt
 1561 ---------+---------+---------+---------+---------+---------+ 1620
      gacgtgaccacaaaacaacacccctcctcctacccctcatcctgtatggtcgaatctaaa

a       L  H  W  C  F  V  V  G  R  R  M  G  S  R  T  Y  Q  L  R  F  -
b        C  T  G  V  L  L  W  G  G  G  W  G  V  G  H  T  S  L  D  F  -
c         A  L  V  F  C  C  G  E  E  D  G  E  *  D  I  P  A  *  I  L -

      taaggttttactgtgagggatgtttgggagatgtaagaaatgttcttgcagttaagggt
 1621 ---------+---------+---------+---------+---------+---------+ 1680
      attccaaaaatgacactccctacaaaccctctacattctttacaagaacgtcaattccca

a       *  G  F  Y  C  E  G  C  L  G  D  V  R  N  V  L  A  V  K  G  -
b        K  V  F  T  V  R  D  V  W  E  M  *  E  M  F  L  Q  L  R  V  -
c         R  F  L  L  *  G  M  F  G  R  C  K  K  C  S  C  S  *  G  L -

      tagtttacaatcagccacattctaggtagggacccacttcaccgtactaaccagggaagc
 1681 ---------+---------+---------+---------+---------+---------+ 1740
      atcaaatgttagtcggtgtaagatccatccctgggtgaagtggcatgattggtcccttcg

a       *  F  T  I  S  H  I  L  G  R  D  P  L  H  R  T  N  Q  G  S  -
b        S  L  Q  S  A  T  F  *  V  G  T  H  F  T  V  L  T  R  E  A  -
c         V  Y  N  Q  P  H  S  R  *  G  P  T  S  P  Y  *  P  G  K  L -

      tgtccctcactgttgaattc
 1741 ---------+---------+ 1760
      acagggagtgacaacttaag

a       C  P  S  L  L  N     -
b        V  P  H  C  *  I     -
c         S  L  T  V  E  F     -
```

Enzymes that do cut:

NONE

Enzymes that do not cut:

NotI

Figure 15-1 BCL2

(Linear) MAP of: hsbcl2a check: 1433 from: 1 to: 5086

RL;HSBCL2A - Human B-cell leukemia/lymphoma 2 (bcl-2) proto-oncogene mRNA
ID HSBCL2A standard; RNA; HUM; 5086 BP.
AC M13994;
NI g179366
DT 19-SEP-1987 (Rel. 13, Created)
DT 16-DEC-1994 (Rel. 42, Last updated, Version 3) . . .

With 1 enzymes: NOTI

```
          gcgcccgcccctccgcgccgcctgcccgcccgcccgccgcgctcccgcccgccgctctcc
        1 ---------+---------+---------+---------+---------+---------+ 60
          cgcgggcggggaggcgcggcggacgggcgggcgggcggcgcgagggcgggcggcgagagg

a         A  P  A  P  P  R  R  L  P  A  R  P  P  R  S  R  P  P  L  S  -
b          R  P  P  L  R  A  A  C  P  P  A  R  R  A  P  A  R  R  S  P  -
c           A  R  P  S  A  P  P  A  R  P  P  A  A  L  P  P  A  A  L  R -

          gtggccccgccgcgctgccgccgccgccgctgccagcgaaggtgccggggctccgggccc
       61 ---------+---------+---------+---------+---------+---------+ 120
          caccggggcggcgcgacggcggcggcggcgacggtcgcttccacggccccgaggcccggg

a         V  A  P  P  R  C  R  R  R  R  C  Q  R  R  C  R  G  S  G  P  -
b          W  P  R  R  A  A  A  A  A  A  A  S  E  G  A  G  A  P  G  P  -
c           G  P  A  A  L  P  P  P  P  L  P  A  K  V  P  G  L  R  A  L -

          tccctgccggcggccgtcagcgctcggagcgaactgcgcgacgggaggtccgggaggcga
      121 ---------+---------+---------+---------+---------+---------+ 180
          agggacggccgccggcagtcgcgagcctcgcttgacgcgctgccctccaggccctccgct

a         S  L  P  A  A  V  S  A  R  S  E  L  R  D  G  R  S  G  R  R  -
b          P  C  R  R  P  S  A  L  G  A  N  C  A  T  G  G  P  G  G  D  -
c           P  A  G  G  R  Q  R  S  E  R  T  A  R  R  E  V  R  E  A  T -

          ccgtagtcgcgccgccgcgcaggaccaggaggaggagaaagggtgcgcagcccggaggcg
      181 ---------+---------+---------+---------+---------+---------+ 240
          ggcatcagcgcggcggcgcgtcctggtcctcctcctctttcccacgcgtcgggcctccgc

a         P  *  S  R  R  R  A  G  P  G  G  G  E  R  V  R  S  P  E  A  -
b          R  S  R  A  A  A  Q  D  Q  E  E  E  K  G  C  A  A  R  R  R  -
c           V  V  A  P  P  R  R  T  R  R  R  R  K  G  A  Q  P  G  G  G -
```

Figure 15-2

```
     gggtgcgccggtggggtgcagcggaagaggggtccaggggggagaacttcgtagcagtc
241 ---------+---------+---------+---------+---------+---------+ 300
     cccacgcggccaccccacgtcgccttctcccccaggtccccccctcttgaagcatcgtcag

a      G  C  A  G  G  V  Q  R  K  R  G  S  R  G  E  N  F  V  A  V  -
b       G  A  P  V  G  C  S  G  R  G  G  P  G  G  R  T  S  *  Q  S  -
c        V  R  R  W  G  A  A  E  E  G  V  Q  G  G  E  L  R  S  S  H -

     atccttttaggaaaagagggaaaaaataaaaccctcccccaccacctccttctccccac
301 ---------+---------+---------+---------+---------+---------+ 360
     taggaaaaatccttttctccctttttatttgggaggggggtggtggaggaagaggggtg

a      I  L  F  R  K  R  G  K  K  *  N  P  P  P  P  P  P  S  P  H  -
b       S  F  L  G  K  E  G  K  N  K  T  L  P  H  H  L  L  L  P  T  -
c        P  F  *  E  K  R  E  K  I  K  P  S  P  T  T  S  F  S  P  P -

     ccctcgccgcaccacacacagcgcgggcttctagcgctcggcaccggcgggccaggcgcg
361 ---------+---------+---------+---------+---------+---------+ 420
     gggagcggcgtggtgtgtgtcgcgcccgaagatcgcgagccgtggccgcccggtccgcgc

a      P  S  P  H  H  T  Q  R  G  L  L  A  L  G  T  G  G  P  G  A  -
b       P  R  R  T  T  H  S  A  G  F  *  R  S  A  P  A  G  Q  A  R  -
c        L  A  A  P  H  T  A  R  A  S  S  A  R  H  R  R  A  R  R  V -

     tcctgccttcatttatccagcagcttttcggaaaatgcatttgctgttcggagtttaatc
421 ---------+---------+---------+---------+---------+---------+ 480
     aggacggaagtaaataggtcgtcgaaaagccttttacgtaaacgacaagcctcaaattag

a      S  C  L  H  L  S  S  S  F  S  E  N  A  F  A  V  R  S  L  I  -
b       P  A  F  I  Y  P  A  A  F  R  K  M  H  L  L  F  G  V  *  S  -
c        L  P  S  F  I  Q  Q  L  F  G  K  C  I  C  C  S  E  F  N  Q -

     agaagacgattcctgcctccgtccccggctccttcatcgtcccatctcccctgtctctct
481 ---------+---------+---------+---------+---------+---------+ 540
     tcttctgctaaggacggaggcaggggccgaggaagtagcagggtagaggggacagagaga

a      R  R  R  F  L  P  P  S  P  A  P  S  S  S  H  L  P  C  L  S  -
b       E  D  D  S  C  L  R  P  R  L  L  H  R  P  I  S  P  V  S  L  -
c        K  T  I  P  A  S  V  P  G  S  F  I  V  P  S  P  L  S  L  S -

     cctggggaggcgtgaagcggtcccgtggatagagattcatgcctgtgtccgcgcgtgtgt
541 ---------+---------+---------+---------+---------+---------+ 600
     ggacccctccgcacttcgccagggcacctatctctaagtacggacacaggcgcgcacaca
```

Figure 15-3

```
a        P  G  E  A  *  S  G  P  V  D  R  D  S  C  L  C  P  R  V  C  -
b         L  G  R  R  E  A  V  P  W  I  E  I  H  A  C  V  R  A  C  V  -
c          W  G  G  V  K  R  S  R  G  *  R  F  M  P  V  S  A  R  V  C -

         gcgcgcgtataaattgccgagaaggggaaaacatcacaggacttctgcgaataccggact
     601 ---------+---------+---------+---------+---------+---------+ 660
         cgcgcgcatatttaacggctcttccccttttgtagtgtcctgaagacgcttatggcctga

a        A  R  V  *  I  A  E  K  G  K  T  S  Q  D  F  C  E  Y  R  T  -
b         R  A  Y  K  L  P  R  R  G  K  H  H  R  T  S  A  N  T  G  L  -
c          A  R  I  N  C  R  E  G  E  N  I  T  G  L  L  R  I  P  D  * -

         gaaaattgtaattcatctgccgccgccgctgccaaaaaaaaactcgagctcttgagatct
     661 ---------+---------+---------+---------+---------+---------+ 720
         cttttaacattaagtagacggcggcggcgacggtttttttttgagctcgagaactctaga

a        E  N  C  N  S  S  A  A  A  A  A  K  K  K  L  E  L  L  R  S  -
b         K  I  V  I  H  L  P  P  P  L  P  K  K  N  S  S  S  *  D  L  -
c          K  L  *  F  I  C  R  R  R  C  Q  K  K  T  R  A  L  E  I  S -

         ccggttgggattcctgcggattgacatttctgtgaagcagaagtctgggaatcgatctgg
     721 ---------+---------+---------+---------+---------+---------+ 780
         ggccaaccctaaggacgcctaactgtaaagacacttcgtcttcagacccttagctagacc

a        P  V  G  I  P  A  D  *  H  F  C  E  A  E  V  W  E  S  I  W  -
b         R  L  G  F  L  R  I  D  I  S  V  K  Q  K  S  G  N  R  S  G  -
c          G  W  D  S  C  G  L  T  F  L  *  S  R  S  L  G  I  D  L  E -

         aaatcctcctaatttttactccctctccccccgactcctgattcattgggaagtttcaaa
     781 ---------+---------+---------+---------+---------+---------+ 840
         tttaggaggattaaaaatgagggagagggggggctgaggactaagtaacccttcaaagttt

a        K  S  S  *  F  L  L  P  L  P  P  T  P  D  S  L  G  S  F  K  -
b         N  P  P  N  F  Y  S  L  S  P  R  L  L  I  H  W  E  V  S  N  -
c          I  L  L  I  F  T  P  S  P  P  D  S  *  F  I  G  K  F  Q  I -

         tcagctataactggagagtgctgaagattgatgggatcgttgccttatgcatttgttttg
     841 ---------+---------+---------+---------+---------+---------+ 900
         agtcgatattgacctctcacgacttctaactaccctagcaacggaatacgtaaacaaaac

a        S  A  I  T  G  E  C  *  R  L  M  G  S  L  P  Y  A  F  V  L  -
b         Q  L  *  L  E  S  A  E  D  *  W  D  R  C  L  M  H  L  F  W  -
c          S  Y  N  W  R  V  L  K  I  D  G  I  V  A  L  C  I  C  F  G -
```

Figure 15-4

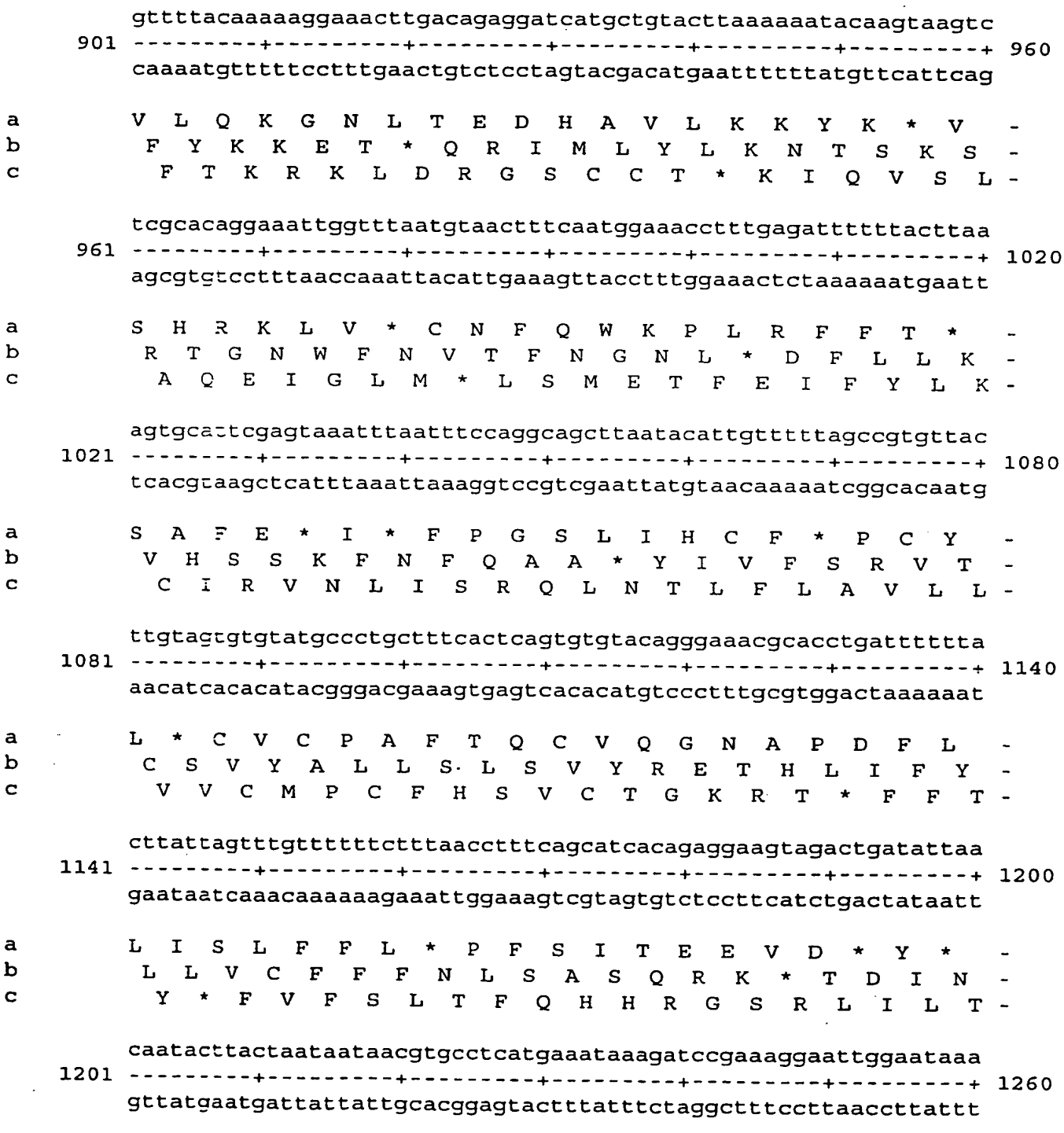
gttttacaaaaaggaaacttgacagaggatcatgctgtacttaaaaaatacaagtaagtc
901 ---------+---------+---------+---------+---------+---------+ 960
caaaatgtttttcctttgaactgtctcctagtacgacatgaatttttatgttcattcag
a V L Q K G N L T E D H A V L K K Y K * V -
b F Y K K E T * Q R I M L Y L K N T S K S -
c F T K R K L D R G S C C T * K I Q V S L -
tcgcacaggaaattggtttaatgtaactttcaatggaaacctttgagatttttacttaa
961 ---------+---------+---------+---------+---------+---------+ 1020
agcgtgtcctttaaccaaattacattgaaagttacctttggaaactctaaaaaatgaatt
a S H R K L V * C N F Q W K P L R F F T * -
b R T G N W F N V T F N G N L * D F L L K -
c A Q E I G L M * L S M E T F E I F Y L K -
agtgcattcgagtaaatttaatttccaggcagcttaatacattgtttttagccgtgttac
1021 ---------+---------+---------+---------+---------+---------+ 1080
tcacgtaagctcatttaaattaaaggtccgtcgaattatgtaacaaaaatcggcacaatg
a S A F E * I * F P G S L I H C F * P C Y -
b V H S S K F N F Q A A * Y I V F S R V T -
c C I R V N L I S R Q L N T L F L A V L L -
ttgtagtgtgtatgccctgctttcactcagtgtgtacagggaaacgcacctgatttttta
1081 ---------+---------+---------+---------+---------+---------+ 1140
aacatcacacatacgggacgaaagtgagtcacacatgtccctttgcgtggactaaaaaat
a L * C V C P A F T Q C V Q G N A P D F L -
b C S V Y A L L S L S V Y R E T H L I F Y -
c V V C M P C F H S V C T G K R T * F F T -
cttattagtttgttttttctttaacctttcagcatcacagaggaagtagactgatattaa
1141 ---------+---------+---------+---------+---------+---------+ 1200
gaataatcaaacaaaaaagaaattggaaagtcgtagtgtctccttcatctgactataatt
a L I S L F F L * P F S I T E E V D * Y * -
b L L V C F F F N L S A S Q R K * T D I N -
c Y * F V F S L T F Q H H R G S R L I L T -
caatacttactaataataacgtgcctcatgaaataaagatccgaaaggaattggaataaa
1201 ---------+---------+---------+---------+---------+---------+ 1260
gttatgaatgattattattgcacggagtactttatttctaggctttccttaaccttattt

Figure 15-5

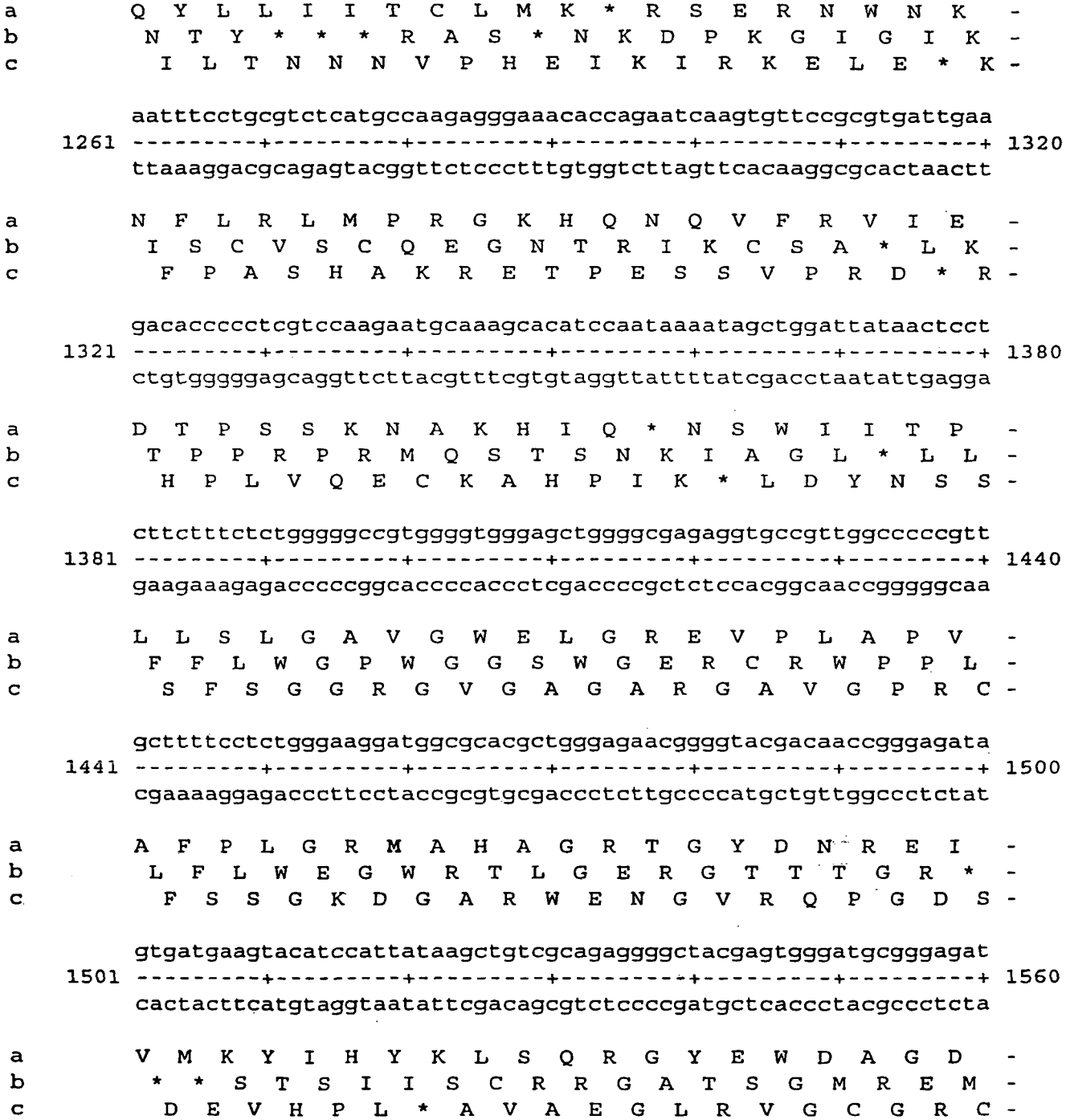
a Q Y L L I I T C L M K * R S E R N W N K -
b N T Y * * * R A S * N K D P K G I G I K -
c I L T N N N V P H E I K I R K E L E * K -
aatttcctgcgtctcatgccaagagggaaacaccagaatcaagtgttccgcgtgattgaa
1261 ttaaaggacgcagagtacggttctcccttgtggtcttagttcacaaggcgcactaactt 1320
a N F L R L M P R G K H Q N Q V F R V I E -
b I S C V S C Q E G N T R I K C S A * L K -
c F P A S H A K R E T P E S S V P R D * R -
gacacccccctcgtccaagaatgcaaagcacatccaataaaatagctggattataactcct
1321 ctgtgggggagcaggttcttacgtttcgtgtaggttattttatcgacctaatattgagga 1380
a D T P S S K N A K H I Q * N S W I I T P -
b T P P R P R M Q S T S N K I A G L * L L -
c H P L V Q E C K A H P I K * L D Y N S S -
cttctttctctgggggccgtggggtgggagctggggcgagaggtgccgttggcccccgtt
1381 gaagaaagagacccccggcaccccaccctcgaccccgctctccacggcaaccgggggcaa 1440
a L L S L G A V G W E L G R E V P L A P V -
b F F L W G P W G G S W G E R C R W P P L -
c S F S G G R G V G A G A R G A V G P R C -
gcttttcctctgggaaggatggcgcacgctgggagaacggggtacgacaaccgggagata
1441 cgaaaaggagacccttcctaccgcgtgcgaccctcttgccccatgctgttggccctctat 1500
a A F P L G R M A H A G R T G Y D N R E I -
b L F L W E G W R T L G E R G T T T G R * -
c F S S G K D G A R W E N G V R Q P G D S -
gtgatgaagtacatccattataagctgtcgcagaggggctacgagtgggatgcgggagat
1501 cactacttcatgtaggtaatattcgacagcgtctccccgatgctcaccctacgccctcta 1560
a V M K Y I H Y K L S Q R G Y E W D A G D -
b * * S T S I I S C R R G A T S G M R E M -
c D E V H P L * A V A E G L R V G C G R C -

Figure 15-6

```
        gtgggcgccgcgcccccgggggccgcccccgcaccgggcatcttctcctcccagcccggg
    1561 ---------+---------+---------+---------+---------+---------+ 1620
        cacccgcggcgcgggggcccccggcgggggcgtggcccgtagaagaggagggtcgggccc

a           V  G  A  A  P  P  G  A  A  P  A  P  G  I  F  S  S  Q  P  G  -
b            W  A  P  R  P  R  G  P  P  P  H  R  A  S  S  P  P  S  P  G  -
c             G  R  R  A  P  G  G  R  P  R  T  G  H  L  L  L  P  A  R  A -

        cacacgccccatccagccgcatcccgcgacccggtcgccaggacctcgccgctgcagacc
    1621 ---------+---------+---------+---------+---------+---------+ 1680
        gtgtgcggggtaggtcggcgtagggcgctgggccagcggtcctggagcggcgacgtctgg

a           H  T  P  H  P  A  A  S  R  D  P  V  A  R  T  S  P  L  Q  T  -
b            T  R  P  I  Q  P  H  P  A  T  R  S  P  G  P  R  R  C  R  P  -
c             H  A  P  S  S  R  I  P  R  P  G  R  Q  D  L  A  A  A  D  P -

        ccggctgcccccggcgccgccgcggggcctgcgctcagcccggtgccacctgtggtccac
    1681 ---------+---------+---------+---------+---------+---------+ 1740
        ggccgacgggggccgcggcggcgccccggacgcgagtcgggccacggtggacaccaggtg

a           P  A  A  P  G  A  A  A  G  P  A  L  S  P  V  P  P  V  V  H  -
b            R  L  P  P  A  P  P  R  G  L  R  S  A  R  C  H  L  W  S  T  -
c             G  C  P  R  R  R  R  G  A  C  A  Q  P  G  A  T  C  G  P  P -

        ctggccctccgccaagccggcgacgacttctcccgccgctaccgcggcgacttcgccgag
    1741 ---------+---------+---------+---------+---------+---------+ 1800
        gaccgggaggcggttcggccgctgctgaagagggcggcgatggcgccgctgaagcggctc

a           L  A  L  R  Q  A  G  D  D  F  S  R  R  Y  R  G  D  F  A  E  -
b            W  P  S  A  K  P  A  T  T  S  P  A  A  T  A  A  T  S  P  R  -
c             G  P  P  P  S  R  R  R  L  L  P  P  L  P  R  R  L  R  R  D -

        atgtccagccagctgcacctgacgcccttcaccgcgcggggacgctttgccacggtggtg
    1801 ---------+---------+---------+---------+---------+---------+ 1860
        tacaggtcggtcgacgtggactgcgggaagtggcgcgcccctgcgaaacggtgccaccac

a           M  S  S  Q  L  H  L  T  P  F  T  A  R  G  R  F  A  T  V  V  -
b            C  P  A  S  C  T  *  R  P  S  P  R  G  D  A  L  P  R  W  W  -
c             V  Q  P  A  A  P  D  A  L  H  R  A  G  T  L  C  H  G  G  G -

        gaggagctcttcagggacggggtgaactgggggaggattgtggccttctttgagttcggt
    1861 ---------+---------+---------+---------+---------+---------+ 1920
        ctcctcgagaagtccctgccccacttgacccccctcctaacaccggaagaaactcaagcca
```

Figure 15-7

```
a        E  E  L  F  R  D  G  V  N  W  G  R  I  V  A  F  F  E  F  G   -
b         R  S  S  S  G  T  G  *  T  G  G  G  L  W  P  S  L  S  S  V  -
c          G  A  L  Q  G  R  G  E  L  G  E  D  C  G  L  L  *  V  R  W -

         ggggtcatgtgtgtggagagcgtcaaccgggagatgtcgcccctggtggacaacatcgcc
    1921 ---------+---------+---------+---------+---------+---------+ 1980
         ccccagtacacacacctctcgcagttggccctctacagcggggaccacctgttgtagcgg

a        G  V  M  C  V  E  S  V  N  R  E  M  S  P  L  V  D  N  I  A   -
b         G  S  C  V  W  R  A  S  T  G  R  C  R  P  W  W  T  T  S  P  -
c          G  H  V  C  G  E  R  Q  P  G  D  V  A  P  G  G  Q  H  R  P -

         ctgtggatgactgagtacctgaaccggcacctgcacacctggatccaggataacggaggc
    1981 ---------+---------+---------+---------+---------+---------+ 2040
         gacacctactgactcatggacttggccgtggacgtgtggacctaggtcctattgcctccg

a        L  W  M  T  E  Y  L  N  R  H  L  H  T  W  I  Q  D  N  G  G   -
b         C  G  *  L  S  T  *  T  G  T  C  T  P  G  S  R  I  T  E  A  -
c          V  D  D  *  V  P  E  P  A  P  A  H  L  D  P  G  *  R  R  L -

         tgggatgcctttgtggaactgtacggccccagcatgcggcctctgtttgatttctcctgg
    2041 ---------+---------+---------+---------+---------+---------+ 2100
         accctacggaaacaccttgacatgccggggtcgtacgccggagacaaactaaagaggacc

a        W  D  A  F  V  E  L  Y  G  P  S  M  R  P  L  F  D  F  S  W   -
b         G  M  P  L  W  N  C  T  A  P  A  C  G  L  C  L  I  S  P  G  -
c          G  C  L  C  G  T  V  R  P  Q  H  A  A  S  V  *  F  L  L  A -

         ctgtctctgaagactctgctcagtttggccctggtgggagcttgcatcaccctgggtgcc
    2101 ---------+---------+---------+---------+---------+---------+ 2160
         gacagagacttctgagacgagtcaaaccgggaccaccctcgaacgtagtgggacccacgg

a        L  S  L  K  T  L  L  S  L  A  L  V  G  A  C  I  T  L  G  A   -
b         C  L  *  R  L  C  S  V  W  P  W  W  E  L  A  S  P  W  V  P  -
c          V  S  E  D  S  A  Q  F  G  P  G  G  S  L  H  H  P  G  C  L -

         tatctgagccacaagtgaagtcaacatgcctgccccaaacaaatatgcaaaaggttcact
    2161 ---------+---------+---------+---------+---------+---------+ 2220
         atagactcggtgttcacttcagttgtacggacggggtttgtttatacgttttccaagtga

a        Y  L  S  H  K  *  S  Q  H  A  C  P  K  Q  I  C  K  R  F  T   -
b         I  *  A  T  S  E  V  N  M  P  A  P  N  K  Y  A  K  G  S  L  -
c          S  E  P  Q  V  K  S  T  C  L  P  Q  T  N  M  Q  K  V  H  * -
```

Figure 15-8

```
        aaagcagtagaaataatatgcattgtcagtgatgtaccatgaaacaaagctgcaggctgt
    2221 ---------+---------+---------+---------+---------+---------+ 2280
        tttcgtcatctttattatacgtaacagtcactacatggtactttgtttcgacgtccgaca

a           K  A  V  E  I  I  C  I  V  S  D  V  P  *  N  K  A  A  G  C  -
b            K  Q  *  K  *  Y  A  L  S  V  M  Y  H  E  T  K  L  Q  A  V  -
c             S  S  R  N  N  M  H  C  Q  *  C  T  M  K  Q  S  C  R  L  F -

        ttaagaaaaaataacacacatataaacatcacacacacagacagacacacacacacacaa
    2281 ---------+---------+---------+---------+---------+---------+ 2340
        aattctttttttattgtgtgtatatttgtagtgtgtgtgtctgtctgtgtgtgtgtgtgtt

a           L  R  K  N  N  T  H  I  N  I  T  H  T  D  R  H  T  H  T  Q  -
b            *  E  K  I  T  H  I  *  T  S  H  T  Q  T  D  T  H  T  H  N  -
c             K  K  K  *  H  T  Y  K  H  H  T  H  R  Q  T  H  T  H  T  T -

        caattaacagtcttcaggcaaaacgtcgaatcagctatttactgccaaagggaaatatca
    2341 ---------+---------+---------+---------+---------+---------+ 2400
        gttaattgtcagaagtccgttttgcagcttagtcgataaatgacggtttccctttatagt

a           Q  L  T  V  F  R  Q  N  V  E  S  A  I  Y  C  Q  R  E  I  S  -
b            N  *  Q  S  S  G  K  T  S  N  Q  L  F  T  A  K  G  K  Y  H  -
c             I  N  S  L  Q  A  K  R  R  I  S  Y  L  L  P  K  G  N  I  I -

        tttatttttacattattaagaaaaaagatttatttatttaagacagtcccatcaaaact
    2401 ---------+---------+---------+---------+---------+---------+ 2460
        aaataaaaaatgtaataattctttttttctaaataaataaattctgtcagggtagttttga

a           F  I  F  Y  I  I  K  K  K  D  L  F  I  *  D  S  P  I  K  T  -
b            L  F  F  T  L  L  R  K  K  I  Y  L  F  K  T  V  P  S  K  L  -
c             Y  F  L  H  Y  *  E  K  R  F  I  Y  L  R  Q  S  H  Q  N  S -

        ccgtctttggaaatccgaccactaattgccaaacaccgcttcgtgtggctccacctggat
    2461 ---------+---------+---------+---------+---------+---------+ 2520
        ggcagaaacctttaggctggtgattaacggtttgtggcgaagcacaccgaggtggaccta

a           P  S  L  E  I  R  P  L  I  A  K  H  R  F  V  W  L  H  L  D  -
b            R  L  W  K  S  D  H  *  L  P  N  T  A  S  C  G  S  T  W  M  -
c             V  F  G  N  P  T  T  N  C  Q  T  P  L  R  V  A  P  P  G  C -

        gttctgtgcctgtaaacatagattcgctttccatgttgttggccggatcaccatctgaag
    2521 ---------+---------+---------+---------+---------+---------+ 2580
        caagacacggacatttgtatctaagcgaaaggtacaacaaccggcctagtggtagacttc
```

Figure 15-9

```
a        V  L  C  L  *  T  *  I  R  F  P  C  C  W  P  D  H  H  L  K  -
b         F  C  A  C  K  H  R  F  A  F  H  V  V  G  R  I  T  I  *  R  -
c          S  V  P  V  N  I  D  S  L  S  M  L  L  A  G  S  P  S  E  E -

         agcagacggatggaaaaaggacctgatcattggggaagctggctttctggctgctggagg
    2581 ---------+---------+---------+---------+---------+---------+ 2640
         tcgtctgcctacctttttcctggactagtaacccccttcgaccgaaagaccgacgacctcc

a        S  R  R  M  E  K  G  P  D  H  W  G  S  W  L  S  G  C  W  R  -
b         A  D  G  W  K  K  D  L  I  I  G  E  A  G  F  L  A  A  G  G  -
c          Q  T  D  G  K  R  T  *  S  L  G  K  L  A  F  W  L  L  E  A -

         ctggggagaaggtgttcattcacttgcatttctttgccctgggggcgtgatattaacaga
    2641 ---------+---------+---------+---------+---------+---------+ 2700
         gacccctcttccacaagtaagtgaacgtaaagaaacgggacccccgcactataattgtct

a        L  G  R  R  C  S  F  T  C  I  S  L  P  W  G  R  D  I  N  R  -
b         W  G  E  G  V  H  S  L  A  F  L  C  P  G  G  V  I  L  T  E  -
c          G  E  K  V  F  I  H  L  H  F  F  A  L  G  A  *  Y  *  Q  R -

         gggagggttcccgtgggggggaagtccatgcctccctggcctgaagaagagactctttgca
    2701 ---------+---------+---------+---------+---------+---------+ 2760
         ccctcccaagggcacccccttcaggtacggagggaccggacttcttctctgagaaacgt

a        G  R  V  P  V  G  G  S  P  C  L  P  G  L  K  K  R  L  F  A  -
b         G  G  F  P  W  G  E  V  H  A  S  L  A  *  R  R  D  S  L  H  -
c          E  G  S  R  G  G  K  S  M  P  P  W  P  E  E  E  T  L  C  I -

         tatgactcacatgatgcatacctggtgggaggaaaagagttgggaacttcagatggacct
    2761 ---------+---------+---------+---------+---------+---------+ 2820
         atactgagtgtactacgtatggaccaccctccttttctcaacccttgaagtctacctgga

a        Y  D  S  H  D  A  Y  L  V  G  G  K  E  L  G  T  S  D  G  P  -
b         M  T  H  M  M  H  T  W  W  E  E  K  S  W  E  L  Q  M  D  L  -
c          *  L  T  *  C  I  P  G  G  R  K  R  V  G  N  F  R  W  T  * -

         agtacccactgagatttccacgccgaaggacagcgatgggaaaaatgcccttaaatcata
    2821 ---------+---------+---------+---------+---------+---------+ 2880
         tcatgggtgactctaaaggtgcggcttcctgtcgctacccttttacgggaatttagtat

a        S  T  H  *  D  F  H  A  E  G  Q  R  W  E  K  C  P  *  I  I  -
b         V  P  T  E  I  S  T  P  K  D  S  D  G  K  N  A  L  K  S  *  -
c          Y  P  L  R  F  P  R  R  R  T  A  M  G  K  M  P  L  N  H  R -
```

Figure 15-10

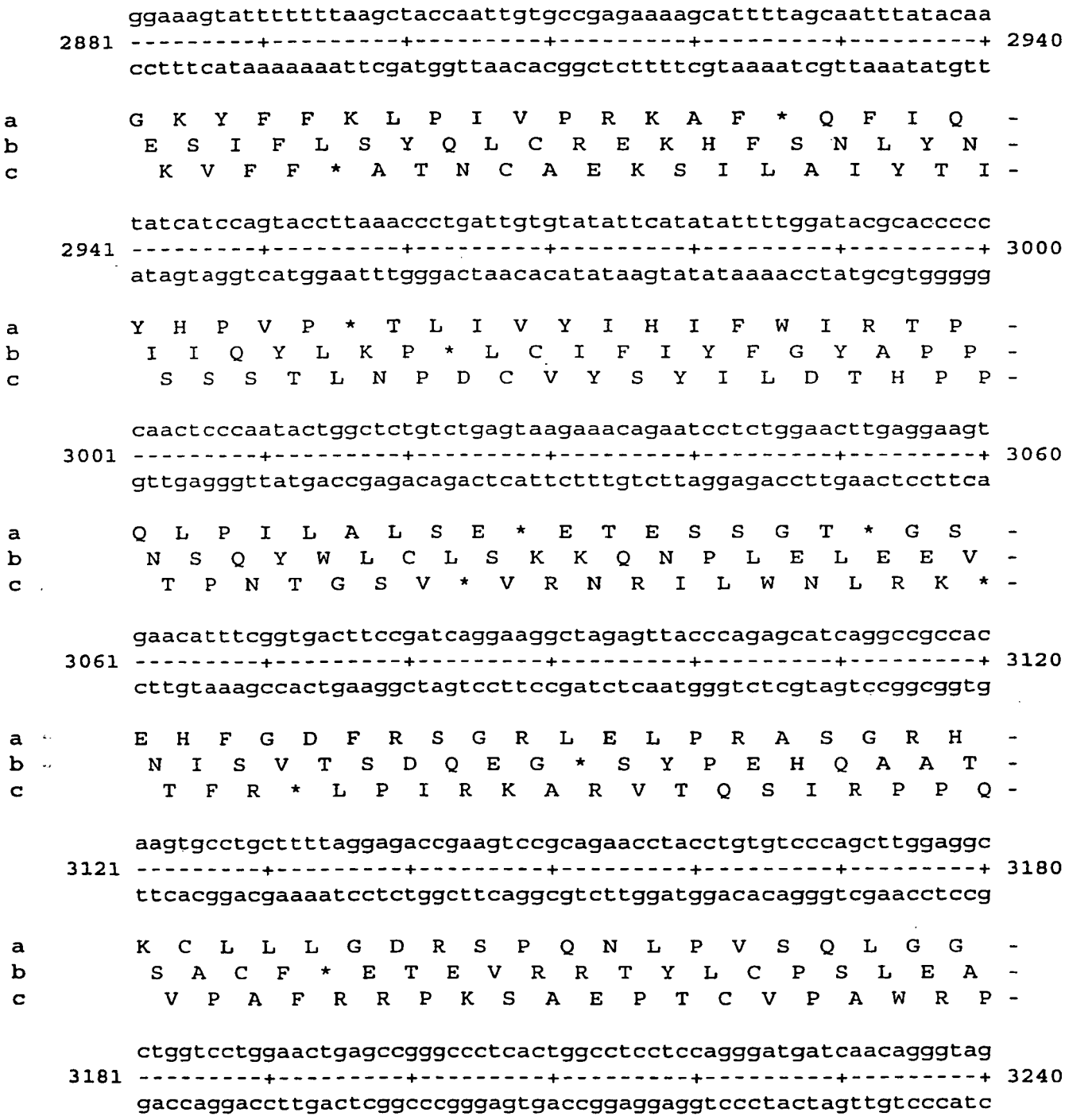
ggaaagtatttttttaagctaccaattgtgccgagaaaagcattttagcaatttatacaa
2881 ---------+---------+---------+---------+---------+---------+ 2940
cctttcataaaaaaattcgatggttaacacggctcttttcgtaaaatcgttaaatatgtt
a G K Y F F K L P I V P R K A F * Q F I Q -
b E S I F L S Y Q L C R E K H F S N L Y N -
c K V F F * A T N C A E K S I L A I Y T I -
tatcatccagtaccttaaaccctgattgtgtatattcatatattttggatacgcaccccc
2941 ---------+---------+---------+---------+---------+---------+ 3000
atagtaggtcatggaatttgggactaacacatataagtatataaaacctatgcgtggggg
a Y H P V P * T L I V Y I H I F W I R T P -
b I I Q Y L K P * L C I F I Y F G Y A P P -
c S S S T L N P D C V Y S Y I L D T H P P -
caactcccaatactggctctgtctgagtaagaaacagaatcctctggaacttgaggaagt
3001 ---------+---------+---------+---------+---------+---------+ 3060
gttgagggttatgaccgagacagactcattctttgtcttaggagaccttgaactccttca
a Q L P I L A L S E * E T E S S G T * G S -
b N S Q Y W L C L S K K Q N P L E L E E V -
c T P N T G S V * V R N R I L W N L R K * -
gaacatttcggtgacttccgatcaggaaggctagagttacccagagcatcaggccgccac
3061 ---------+---------+---------+---------+---------+---------+ 3120
cttgtaaagccactgaaggctagtccttccgatctcaatgggtctcgtagtccggcggtg
a E H F G D F R S G R L E L P R A S G R H -
b N I S V T S D Q E G * S Y P E H Q A A T -
c T F R * L P I R K A R V T Q S I R P P Q -
aagtgcctgcttttaggagaccgaagtccgcagaacctacctgtgtcccagcttggaggc
3121 ---------+---------+---------+---------+---------+---------+ 3180
ttcacggacgaaaatcctctggcttcaggcgtcttggatggacacagggtcgaacctccg
a K C L L L G D R S P Q N L P V S Q L G G -
b S A C F * E T E V R R T Y L C P S L E A -
c V P A F R R P K S A E P T C V P A W R P -
ctggtcctggaactgagccgggccctcactggcctcctccagggatgatcaacagggtag
3181 ---------+---------+---------+---------+---------+---------+ 3240
gaccaggaccttgactcggcccgggagtgaccggaggaggtccctactagttgtcccatc

Figure 15-11

```
a        L  V  L  E  L  S  R  A  L  T  G  L  L  Q  G  *  S  T  G  *   -
b         W  S  W  N  *  A  G  P  S  L  A  S  S  R  D  D  Q  Q  G  S  -
c          G  P  G  T  E  P  G  P  H  W  P  P  P  G  M  I  N  R  V  V -

         tgtggtctccgaatgtctggaagctgatggatggagctcagaattccactgtcaagaaag
    3241 ---------+---------+---------+---------+---------+---------+ 3300
         acaccagaggcttacagaccttcgactacctacctcgagtcttaaggtgacagttctttc

a        C  G  L  R  M  S  G  S  *  W  M  E  L  R  I  P  L  S  R  K   -
b         V  V  S  E  C  L  E  A  D  G  W  S  S  E  F  H  C  Q  E  R  -
c          W  S  P  N  V  W  K  L  M  D  G  A  Q  N  S  T  V  K  K  E -

         agcagtagaggggtgtggctgggcctgtcaccctggggccctccaggtaggcccgttttc
    3301 ---------+---------+---------+---------+---------+---------+ 3360
         tcgtcatctccccacaccgacccggacagtgggaccccgggaggtccatccgggcaaaag

a        S  S  R  G  V  W  L  G  L  S  P  W  G  P  P  G  R  P  V  F   -
b         A  V  E  G  C  G  W  A  C  H  P  G  A  L  Q  V  G  P  F  S  -
c          Q  *  R  G  V  A  G  P  V  T  L  G  P  S  R  *  A  R  F  H -

         acgtggagcataggagccacgacccttcttaagacatgtatcactgtagagggaaggaac
    3361 ---------+---------+---------+---------+---------+---------+ 3420
         tgcacctcgtatcctcggtgctgggaagaattctgtacatagtgacatctcccttccttg

a        T  W  S  I  G  A  T  T  L  L  K  T  C  I  T  V  E  G  R  N   -
b         R  G  A  *  E  P  R  P  F  L  R  H  V  S  L  *  R  E  G  T  -
c          V  E  H  R  S  H  D  P  S  *  D  M  Y  H  C  R  G  K  E  Q -

         agaggccctgggccttcctatcagaaggacatggtgaaggctgggaacgtgaggagaggc
    3421 ---------+---------+---------+---------+---------+---------+ 3480
         tctccgggacccggaaggatagtcttcctgtaccacttccgacccttgcactcctctccg

a        R  G  P  G  P  S  Y  Q  K  D  M  V  K  A  G  N  V  R  R  G   -
b         E  A  L  G  L  P  I  R  R  T  W  *  R  L  G  T  *  G  E  A  -
c          R  P  W  A  F  L  S  E  G  H  G  E  G  W  E  R  E  E  R  Q -

         aatggccacggcccattttggctgtagcacatggcacgttggctgtgtggccttggccac
    3481 ---------+---------+---------+---------+---------+---------+ 3540
         ttaccggtgccgggtaaaaccgacatcgtgtaccgtgcaaccgacacaccggaaccggtg

a        N  G  H  G  P  F  W  L  *  H  M  A  R  W  L  C  G  L  G  H   -
b         M  A  T  A  H  F  G  C  S  T  W  H  V  G  C  V  A  L  A  T  -
c          W  P  R  P  I  L  A  V  A  H  G  T  L  A  V  W  P  W  P  P -
```

Figure 15-12

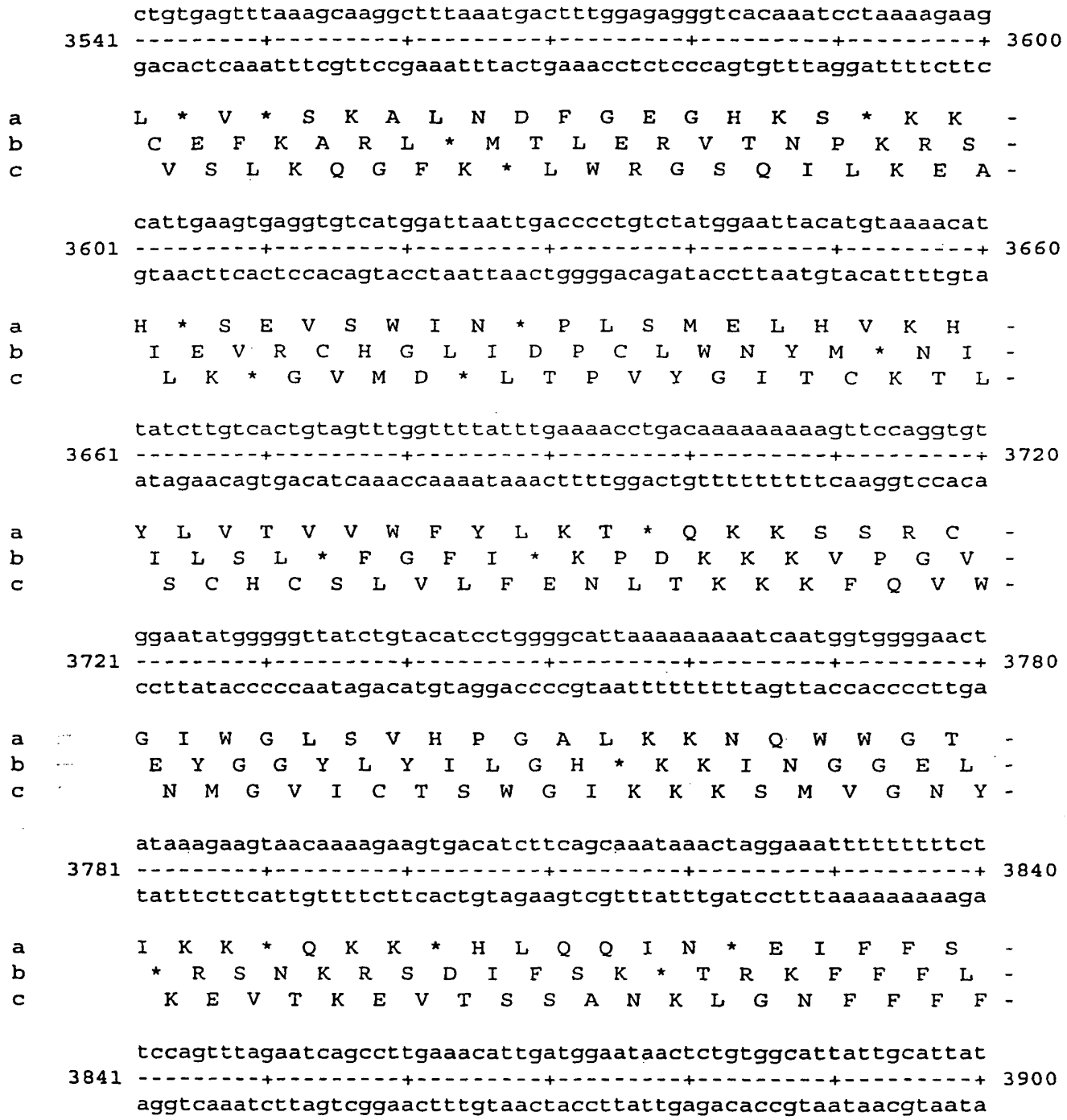
ctgtgagtttaaagcaaggctttaaatgactttggagagggtcacaaatcctaaaagaag
3541 ---------+---------+---------+---------+---------+---------+ 3600
gacactcaaatttcgttccgaaatttactgaaacctctcccagtgtttaggattttcttc
a L * V * S K A L N D F G E G H K S * K K -
b C E F K A R L * M T L E R V T N P K R S -
c V S L K Q G F K * L W R G S Q I L K E A -
cattgaagtgaggtgtcatggattaattgacccctgtctatggaattacatgtaaaacat
3601 ---------+---------+---------+---------+---------+---------+ 3660
gtaacttcactccacagtacctaattaactggggacagataccttaatgtacattttgta
a H * S E V S W I N * P L S M E L H V K H -
b I E V R C H G L I D P C L W N Y M * N I -
c L K * G V M D * L T P V Y G I T C K T L -
tatcttgtcactgtagtttggttttatttgaaaacctgacaaaaaaaaagttccaggtgt
3661 ---------+---------+---------+---------+---------+---------+ 3720
atagaacagtgacatcaaaccaaaataaacttttggactgtttttttttcaaggtccaca
a Y L V T V V W F Y L K T * Q K K S S R C -
b I L S L * F G F I * K P D K K K V P G V -
c S C H C S L V L F E N L T K K K F Q V W -
ggaatatggggttatctgtacatcctggggcattaaaaaaaaatcaatggtggggaact
3721 ---------+---------+---------+---------+---------+---------+ 3780
ccttatacccccaatagacatgtaggaccccgtaatttttttttagttaccacccccttga
a G I W G L S V H P G A L K K N Q W W G T -
b E Y G G Y L Y I L G H * K K I N G G E L -
c N M G V I C T S W G I K K K S M V G N Y -
ataaagaagtaacaaaagaagtgacatcttcagcaaataaactaggaaatttttttttct
3781 ---------+---------+---------+---------+---------+---------+ 3840
tatttcttcattgttttcttcactgtagaagtcgtttatttgatcctttaaaaaaaaaga
a I K K * Q K K * H L Q Q I N * E I F F S -
b * R S N K R S D I F S K * T R K F F F L -
c K E V T K E V T S S A N K L G N F F F F -
tccagtttagaatcagccttgaaacattgatggaataactctgtggcattattgcattat
3841 ---------+---------+---------+---------+---------+---------+ 3900
aggtcaaatcttagtcggaactttgtaactaccttattgagacaccgtaataacgtaata

Figure 15-13

```
a        S  S  L  E  S  A  L  K  H  *  W  N  N  S  V  A  L  L  H  Y  -
b         P  V  *  N  Q  P  *  N  I  D  G  I  T  L  W  H  Y  C  I  I  -
c          Q  F  R  I  S  L  E  T  L  M  E  *  L  C  G  I  I  A  L  Y -

         ataccatttatctgtattaactttggaatgtactctgttcaatgtttaatgctgtggttg
    3901 ---------+---------+---------+---------+---------+---------+ 3960
         tatggtaaatagacataattgaaaccttacatgagacaagttacaaattacgacaccaac

a        I  P  F  I  C  I  N  F  G  M  Y  S  V  Q  C  L  M  L  W  L  -
b         Y  H  L  S  V  L  T  L  E  C  T  L  F  N  V  *  C  C  G  *  -
c          T  I  Y  L  Y  *  L  W  N  V  L  C  S  M  F  N  A  V  V  D -

         atatttcgaaagctgctttaaaaaaatacatgcatctcagcgtttttttgtttttaattg
    3961 ---------+---------+---------+---------+---------+---------+ 4020
         tataaagctttcgacgaaatttttttatgtacgtagagtcgcaaaaaaacaaaaattaac

a        I  F  R  K  L  L  *  K  N  T  C  I  S  A  F  F  C  F  *  L  -
b         Y  F  E  S  C  F  K  K  I  H  A  S  Q  R  F  F  V  F  N  C  -
c          I  S  K  A  A  L  K  K  Y  M  H  L  S  V  F  L  F  L  I  V -

         tatttagttatggcctatacactatttgtgagcaaaggtgatcgttttctgtttgagatt
    4021 ---------+---------+---------+---------+---------+---------+ 4080
         ataaatcaataccggatatgtgataaacactcgtttccactagcaaaagacaaactctaa

a        Y  L  V  M  A  Y  T  L  F  V  S  K  G  D  R  F  L  F  E  I  -
b         I  *  L  W  P  I  H  Y  L  *  A  K  V  I  V  F  C  L  R  F  -
c          F  S  Y  G  L  Y  T  I  C  E  Q  R  *  S  F  S  V  *  D  F -

         tttatctcttgattcttcaaaagcattctgagaaggtgagataagccctgagtctcagct
    4081 ---------+---------+---------+---------+---------+---------+ 4140
         aaatagagaactaagaagttttcgtaagactcttccactctattcgggactcagagtcga

a        F  I  S  *  F  F  K  S  I  L  R  R  *  D  K  P  *  V  S  A  -
b         L  S  L  D  S  S  K  A  F  *  E  G  E  I  S  P  E  S  Q  L  -
c          Y  L  L  I  L  Q  K  H  S  E  K  V  R  *  A  L  S  L  S  Y -

         acctaagaaaaacctggatgtcactggccactgaggagctttgtttcaaccaagtcatgt
    4141 ---------+---------+---------+---------+---------+---------+ 4200
         tggattctttttggacctacagtgaccggtgactcctcgaaacaaagttggttcagtaca

a        T  *  E  K  P  G  C  H  W  P  L  R  S  F  V  S  T  K  S  C  -
b         P  K  K  N  L  D  V  T  G  H  *  G  A  L  F  Q  P  S  H  V  -
c          L  R  K  T  W  M  S  L  A  T  E  E  L  C  F  N  Q  V  M  C -
```

Figure 15-14

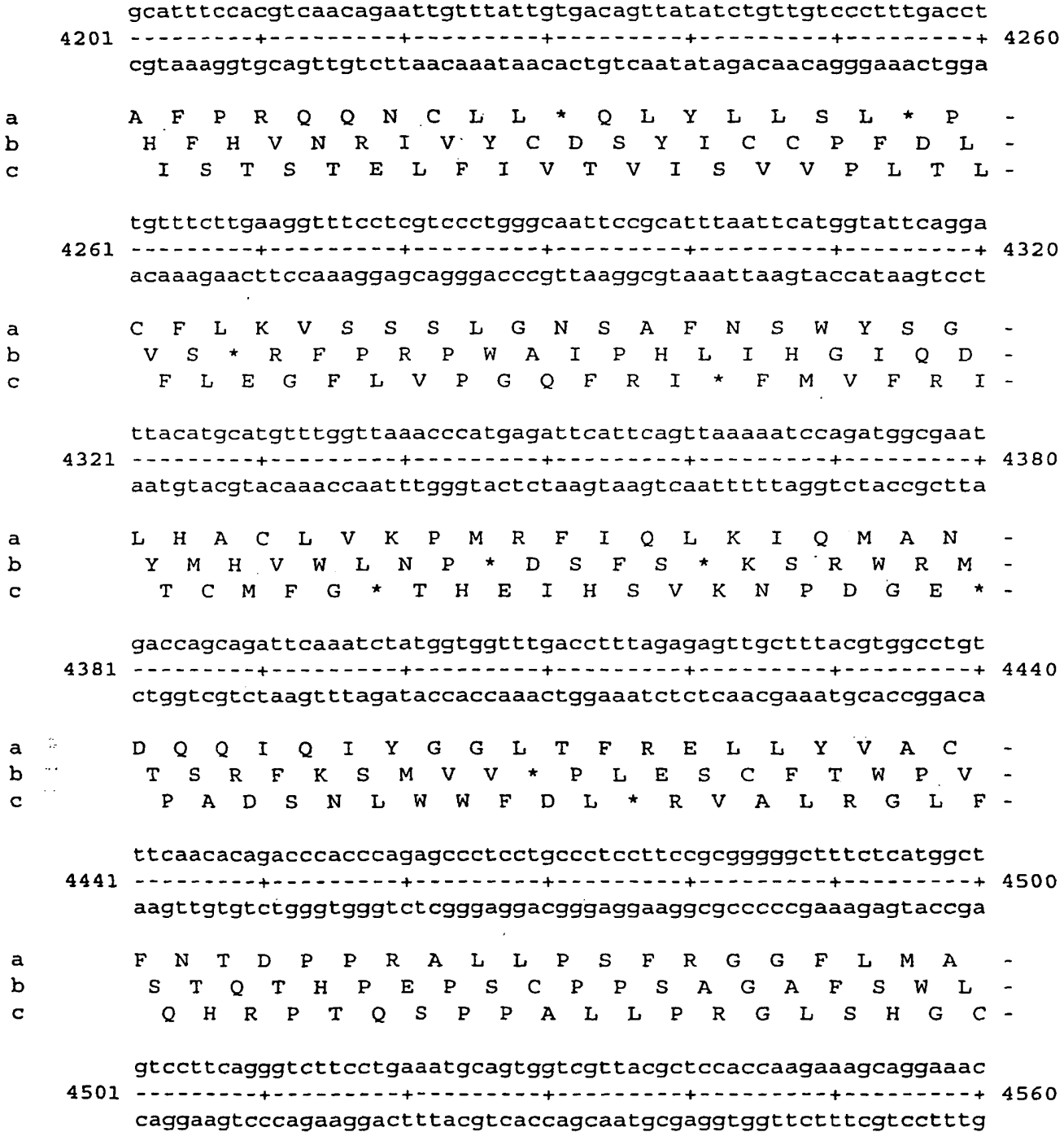
gcatttccacgtcaacagaattgtttattgtgacagttatatctgttgtccctttgacct
4201 4260
cgtaaaggtgcagttgtcttaacaaataacactgtcaatatagacaacagggaaactgga
a A F P R Q Q N C L L * Q L Y L L S L * P -
b H F H V N R I V Y C D S Y I C C P F D L -
c I S T S T E L F I V T V I S V V P L T L -
tgtttcttgaaggtttcctcgtccctgggcaattccgcatttaattcatggtattcagga
4261 4320
acaaagaacttccaaaggagcagggacccgttaaggcgtaaattaagtaccataagtcct
a C F L K V S S S L G N S A F N S W Y S G -
b V S * R F P R P W A I P H L I H G I Q D -
c F L E G F L V P G Q F R I * F M V F R I -
ttacatgcatgtttggttaaacccatgagattcattcagttaaaaatccagatggcgaat
4321 4380
aatgtacgtacaaaccaatttgggtactctaagtaagtcaatttttaggtctaccgctta
a L H A C L V K P M R F I Q L K I Q M A N -
b Y M H V W L N P * D S F S * K S R W R M -
c T C M F G * T H E I H S V K N P D G E * -
gaccagcagattcaaatctatggtggtttgacctttagagagttgctttacgtggcctgt
4381 4440
ctggtcgtctaagtttagataccaccaaactggaaatctctcaacgaaatgcaccggaca
a D Q Q I Q I Y G G L T F R E L L Y V A C -
b T S R F K S M V V * P L E S C F T W P V -
c P A D S N L W W F D L * R V A L R G L F -
ttcaacacagacccacccagagccctcctgccctccttccgcgggggctttctcatggct
4441 4500
aagttgtgtctgggtgggtctcgggaggacgggaggaaggcgcccccgaaagagtaccga
a F N T D P P R A L L P S F R G G F L M A -
b S T Q T H P E P S C P P S A G A F S W L -
c Q H R P T Q S P P A L L P R G L S H G C -
gtccttcagggtcttcctgaaatgcagtggtcgttacgctccaccaagaaagcaggaaac
4501 4560
caggaagtcccagaaggactttacgtcaccagcaatgcgaggtggttctttcgtcctttg

Figure 15-15

```
a        V  L  Q  G  L  P  E  M  Q  W  S  L  R  S  T  K  K  A  G  N  -
b         S  F  R  V  F  L  K  C  S  G  R  Y  A  P  P  R  K  Q  E  T  -
c          P  S  G  S  S  *  N  A  V  V  V  T  L  H  Q  E  S  R  K  P -

         ctgtggtatgaagccagacctccccggcgggcctcagggaacagaatgatcagacctttg
    4561 ---------+---------+---------+---------+---------+---------+---------+ 4620
         gacaccatacttcggtctggaggggccgcccggagtcccttgtcttactagtctggaaac

a        L  W  Y  E  A  R  P  P  R  R  A  S  G  N  R  M  I  R  P  L  -
b         C  G  M  K  P  D  L  P  G  G  P  Q  G  T  E  *  S  D  L  *  -
c          V  V  *  S  Q  T  S  P  A  G  L  R  E  Q  N  D  Q  T  F  E -

         aatgattctaatttttaagcaaaatattattttatgaaaggtttacattgtcaaagtgat
    4621 ---------+---------+---------+---------+---------+---------+---------+ 4680
         ttactaagattaaaaattcgttttataataaaatactttccaaatgtaacagtttcacta

a        N  D  S  N  F  *  A  K  Y  Y  F  M  K  G  L  H  C  Q  S  D  -
b         M  I  L  I  F  K  Q  N  I  I  L  *  K  V  Y  I  V  K  V  M  -
c          *  F  *  F  L  S  K  I  L  F  Y  E  R  F  T  L  S  K  *  * -

         gaatatggaatatccaatcctgtgctgctatcctgccaaaatcattttaatggagtcagt
    4681 ---------+---------+---------+---------+---------+---------+---------+ 4740
         cttataccttataggttaggacacgacgataggacggttttagtaaaattacctcagtca

a        E  Y  G  I  S  N  P  V  L  L  S  C  Q  N  H  F  N  G  V  S  -
b         N  M  E  Y  P  I  L  C  C  Y  P  A  K  I  I  L  M  E  S  V  -
c          I  W  N  I  Q  S  C  A  A  I  L  P  K  S  F  *  W  S  Q  F -

         ttgcagtatgctccacgtggtaagatcctccaagctgctttagaagtaacaatgaagaac
    4741 ---------+---------+---------+---------+---------+---------+---------+ 4800
         aacgtcatacgaggtgcaccattctaggaggttcgacgaaatcttcattgttacttcttg

a        L  Q  Y  A  P  R  G  K  I  L  Q  A  A  L  E  V  T  M  K  N  -
b         C  S  M  L  H  V  V  R  S  S  K  L  L  *  K  *  Q  *  R  T  -
c          A  V  C  S  T  W  *  D  P  P  S  C  F  R  S  N  N  E  E  R -

         gtggacgttttaatataaagcctgttttgtcttttgttgttgttcaaacgggattcaca
    4801 ---------+---------+---------+---------+---------+---------+---------+ 4860
         cacctgcaaaaattatatttcggacaaaacagaaaacaacaacaagtttgccctaagtgt

a        V  D  V  F  N  I  K  P  V  L  S  F  V  V  V  Q  T  G  F  T  -
b         W  T  F  L  I  *  S  L  F  C  L  L  L  L  F  K  R  D  S  Q  -
c          G  R  F  *  Y  K  A  C  F  V  F  C  C  C  S  N  G  I  H  R -
```

Figure 15-16

```
         gagtatttgaaaaatgtatatatattaagaggtcacgggggctaattgctagctggctgc
    4861 ---------+---------+---------+---------+---------+---------+ 4920
         ctcataaacttttacatatatataattctccagtgcccccgattaacgatcgaccgacg

a        E  Y  L  K  N  V  Y  I  L  R  G  H  G  G  *  L  L  A  G  C  -
b         S  I  *  K  M  Y  I  Y  *  E  V  T  G  A  N  C  *  L  A  A  -
c          V  F  E  K  C  I  Y  I  K  R  S  R  G  L  I  A  S  W  L  P -

         cttttgctgtggggttttgttacctggttttaataacagtaaatgtgcccagcctcttgg
    4921 ---------+---------+---------+---------+---------+---------+ 4980
         gaaaacgacaccccaaaacaatggaccaaaattattgtcatttacacgggtcggagaacc

a        L  L  L  W  G  F  V  T  W  F  *  *  Q  *  M  C  P  A  S  W   -
b         F  C  C  G  V  L  L  P  G  F  N  N  S  K  C  A  Q  P  L  G  -
c          F  A  V  G  F  C  Y  L  V  L  I  T  V  N  V  P  S  L  L  A -

         ccccagaactgtacagtattgtggctgcacttgctctaagagtagttgatgttgcatttt
    4981 ---------+---------+---------+---------+---------+---------+ 5040
         ggggtcttgacatgtcataacaccgacgtgaacgagattctcatcaactacaacgtaaaa

a        P  Q  N  C  T  V  L  W  L  H  L  L  *  E  *  L  M  L  H  F   -
b         P  R  T  V  Q  Y  C  G  C  T  C  S  K  S  S  *  C  C  I  F  -
c          P  E  L  Y  S  I  V  A  A  L  A  L  R  V  V  D  V  A  F  S -

         ccttattgttaaaaacatgttagaagcaatgaatgtatataaaagc
    5041 ---------+---------+---------+---------+------ 5086
         ggaataacaatttttgtacaatcttcgttacttacatatattttcg

a        P  Y  C  *  K  H  V  R  S  N  E  C  I  *  K    -
b         L  I  V  K  N  M  L  E  A  M  N  V  Y  K  S   -
c          L  L  L  K  T  C  *  K  Q  *  M  Y  I  K     -
```

Enzymes that do cut:

NONE

Enzymes that do not cut:

NotI

Figure 16-1 Semaphorin III

(Linear) MAP of: hshsem check: 7721 from: 1 to: 2530

RL;HSHSEM - Homo sapiens semaphorin-III (Hsema-I) mRNA, complete cds.
ID HSHSEM standard; RNA; HUM; 2530 BP.
AC L26081;
NI g799328
DT 09-MAR-1994 (Rel. 38, Created)
DT 12-MAY-1995 (Rel. 43, Last updated, Version 3) . . .

With 1 enzymes: NOTI

```
          ggaattccctgcagcatgggctggttaactaggattgtctgtcttttctggggagtatta
        1 ---------+---------+---------+---------+---------+---------+ 60
          ccttaagggacgtcgtacccgaccaattgatcctaacagacagaaaagacccctcataat

a         G  I  P  C  S  M  G  W  L  T  R  I  V  C  L  F  W  G  V  L  -
b          E  F  P  A  A  W  A  G  *  L  G  L  S  V  F  S  G  E  Y  Y  -
c           N  S  L  Q  H  G  L  V  N  *  D  C  L  S  F  L  G  S  I  T -

          cttacagcaagagcaaactatcagaatgggaagaacaatgtgccaaggctgaaattatcc
       61 ---------+---------+---------+---------+---------+---------+ 120
          gaatgtcgttctcgtttgatagtcttacccttcttgttacacggttccgactttaatagg

a         L  T  A  R  A  N  Y  Q  N  G  K  N  N  V  P  R  L  K  L  S  -
b          L  Q  Q  E  Q  T  I  R  M  G  R  T  M  C  Q  G  *  N  Y  P  -
c           Y  S  K  S  K  L  S  E  W  E  E  Q  C  A  K  A  E  I  I  L -

          tacaaagaaatgttggaatccaacaatgtgatcactttcaatggcttggccaacagctcc
      121 ---------+---------+---------+---------+---------+---------+ 180
          atgtttctttacaaccttaggttgttacactagtgaaagttaccgaaccggttgtcgagg

a         Y  K  E  M  L  E  S  N  N  V  I  T  F  N  G  L  A  N  S  S  -
b          T  K  K  C  W  N  P  T  M  *  S  L  S  M  A  W  P  T  A  P  -
c           Q  R  N  V  G  I  Q  Q  C  D  H  F  Q  W  L  G  Q  Q  L  Q -

          agttatcataccttccttttggatgaggaacggagtaggctgtatgttggagcaaaggat
      181 ---------+---------+---------+---------+---------+---------+ 240
          tcaatagtatggaaggaaaacctactccttgcctcatccgacatacaacctcgtttccta

a         S  Y  H  T  F  L  L  D  E  E  R  S  R  L  Y  V  G  A  K  D  -
b          V  I  I  P  S  F  W  M  R  N  G  V  G  C  M  L  E  Q  R  I  -
c           L  S  Y  L  P  F  G  *  G  T  E  *  A  V  C  W  S  K  G  S -
```

Figure 16-2

```
      cacatattttcattcgacctggttaatatcaaggattttcaaaagattgtgtggccagta
  241 ---------+---------+---------+---------+---------+---------+ 300
      gtgtataaaagtaagctggaccaattatagttcctaaaagttttctaacacaccggtcat

a         H  I  F  S  F  D  L  V  N  I  K  D  F  Q  K  I  V  W  P  V  -
b          T  Y  F  H  S  T  W  L  I  S  R  I  F  K  R  L  C  G  Q  Y  -
c           H  I  F  I  R  P  G  *  Y  Q  G  F  S  K  D  C  V  A  S  I -

      tcttacaccagaagagatgaatgcaagtgggctggaaaagacatcctgaaagaatgtgct
  301 ---------+---------+---------+---------+---------+---------+ 360
      agaatgtggtcttctctacttacgttcacccgacctttttctgtaggactttcttacacga

a         S  Y  T  R  R  D  E  C  K  W  A  G  K  D  I  L  K  E  C  A  -
b          L  T  P  E  E  M  N  A  S  G  L  E  K  T  S  *  K  N  V  L  -
c           L  H  Q  K  R  *  M  Q  V  G  W  K  R  H  P  E  R  M  C  * -

      aatttcatcaaggtacttaaggcatataatcagactcacttgtacgcctgtggaacgggg
  361 ---------+---------+---------+---------+---------+---------+ 420
      ttaaagtagttccatgaattccgtatattagtctgagtgaacatgcggacaccttgcccc

a         N  F  I  K  V  L  K  A  Y  N  Q  T  H  L  Y  A  C  G  T  G  -
b          I  S  S  R  Y  L  R  H  I  I  R  L  T  C  T  P  V  E  R  G  -
c           F  H  Q  G  T  *  G  I  *  S  D  S  L  V  R  L  W  N  G  G -

      gcttttcatccaatttgcacctacattgaaattggacatcatcctgaggacaatattttt
  421 ---------+---------+---------+---------+---------+---------+ 480
      cgaaaagtaggttaaacgtggatgtaactttaacctgtagtaggactcctgttataaaaa

a         A  F  H  P  I  C  T  Y  I  E  I  G  H  H  P  E  D  N  I  F  -
b          L  F  I  Q  F  A  P  T  L  K  L  D  I  I  L  R  T  I  F  L  -
c           F  S  S  N  L  H  L  H  *  N  W  T  S  S  *  G  Q  Y  F  * -

      aagctggagaactcacattttgaaaacggccgtgggaagagtccatatgaccctaagctg
  481 ---------+---------+---------+---------+---------+---------+ 540
      ttcgacctcttgagtgtaaaacttttgccggcacccttctcaggtatactgggattcgac

a         K  L  E  N  S  H  F  E  N  G  R  G  K  S  P  Y  D  P  K  L  -
b          S  W  R  T  H  I  L  K  T  A  V  G  R  V  H  M  T  L  S  C  -
c           A  G  E  L  T  F  *  K  R  P  W  E  E  S  I  *  P  *  A  A -

      ctgacagcatccctttttaatagatggagaattatactctggaactgcagctgattttatg
  541 ---------+---------+---------+---------+---------+---------+ 600
      gactgtcgtagggaaaattatctacctcttaatatgagaccttgacgtcgactaaaatac
```

Figure 16-3

```
a        L  T  A  S  L  L  I  D  G  E  L  Y  S  G  T  A  A  D  F  M  -
b         *  Q  H  P  F  *  *  M  E  N  Y  T  L  E  L  Q  L  I  L  W  -
c          D  S  I  P  F  N  R  W  R  I  I  L  W  N  C  S  *  F  Y  G -

         gggcgagactttgctatcttccgaactcttgggcaccaccacccaatcaggacagagcag
     601 ---------+---------+---------+---------+---------+---------+---------+ 660
         cccgctctgaaacgatagaaggcttgagaacccgtggtggtgggttagtcctgtctcgtc

a        G  R  D  F  A  I  F  R  T  L  G  H  H  H  P  I  R  T  E  Q  -
b         G  E  T  L  L  S  S  E  L  L  G  T  T  T  Q  S  G  Q  S  S  -
c          A  R  L  C  Y  L  P  N  S  W  A  P  P  P  N  Q  D  R  A  A -

         catgattccaggtggctcaatgatccaaagttcattagtgcccacctcatctcagagagt
     661 ---------+---------+---------+---------+---------+---------+---------+ 720
         gtactaaggtccaccgagttactaggtttcaagtaatcacgggtggagtagagtctctca

a        H  D  S  R  W  L  N  D  P  K  F  I  S  A  H  L  I  S  E  S  -
b         M  I  P  G  G  S  M  I  Q  S  S  L  V  P  T  S  S  Q  R  V  -
c          *  F  Q  V  A  Q  *  S  K  V  H  *  C  P  P  H  L  R  E  * -

         gacaatcctgaagatgacaaagtatacttttcttccgtgaaaatgcaatagatggagaa
     721 ---------+---------+---------+---------+---------+---------+---------+ 780
         ctgttaggacttctactgtttcatatgaaaaagaaggcacttttacgttatctacctctt

a        D  N  P  E  D  D  K  V  Y  F  F  F  R  E  N  A  I  D  G  E  -
b         T  I  L  K  M  T  K  Y  T  F  S  S  V  K  M  Q  *  M  E  N  -
c          Q  S  *  R  *  Q  S  I  L  F  L  P  *  K  C  N  R  W  R  T -

         cactctggaaaagctactcacgctagaataggtcagatatgcaagaatgactttggaggg
     781 ---------+---------+---------+---------+---------+---------+---------+ 840
         gtgagacctttcgatgagtgcgatcttatccagtctatacgttcttactgaaacctccc

a        H  S  G  K  A  T  H  A  R  I  G  Q  I  C  K  N  D  F  G  G  -
b         T  L  E  K  L  L  T  L  E  *  V  R  Y  A  R  M  T  L  E  G  -
c          L  W  K  S  Y  S  R  *  N  R  S  D  M  Q  E  *  L  W  R  A -

         cacagaagtctggtgaataaatggacaacattcctcaaagctcgtctgatttgctcagtg
     841 ---------+---------+---------+---------+---------+---------+---------+ 900
         gtgtcttcagaccacttatttacctgttgtaaggagtttcgagcagactaaacgagtcac

a        H  R  S  L  V  N  K  W  T  T  F  L  K  A  R  L  I  C  S  V  -
b         T  E  V  W  *  I  N  G  Q  H  S  S  K  L  V  *  F  A  Q  C  -
c          Q  K  S  G  E  *  M  D  N  I  P  Q  S  S  S  D  L  L  S  A -
```

Figure 16-4

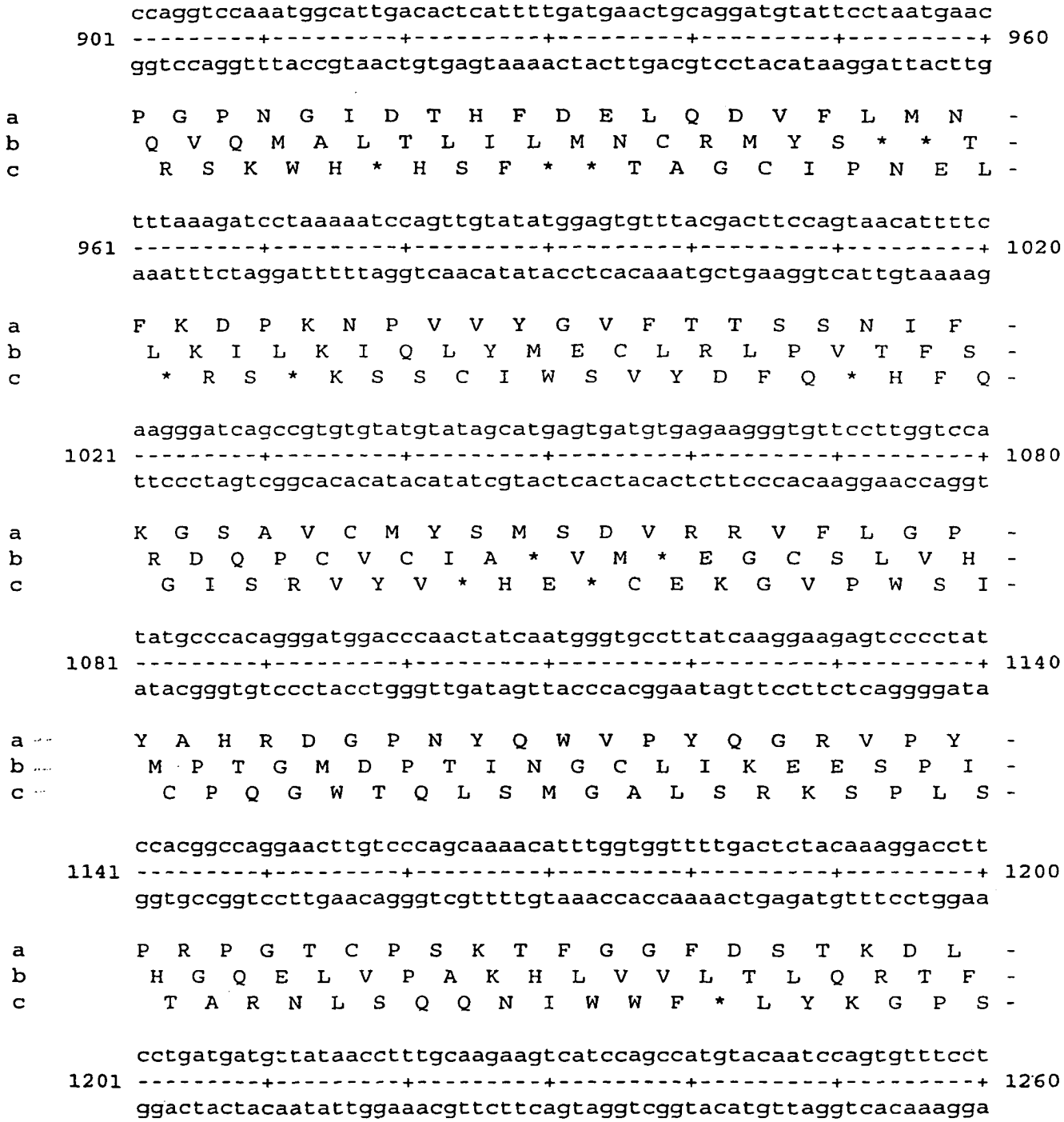
ccaggtccaaatggcattgacactcattttgatgaactgcaggatgtattcctaatgaac
901 ---------+---------+---------+---------+---------+---------+ 960
ggtccaggtttaccgtaactgtgagtaaaactacttgacgtcctacataaggattacttg
a P G P N G I D T H F D E L Q D V F L M N -
b Q V Q M A L T L I L M N C R M Y S * * T -
c R S K W H * H S F * * T A G C I P N E L -
tttaaagatcctaaaaatccagttgtatatggagtgtttacgacttccagtaacattttc
961 ---------+---------+---------+---------+---------+---------+ 1020
aaatttctaggatttttaggtcaacatatacctcacaaatgctgaaggtcattgtaaaag
a F K D P K N P V V Y G V F T T S S N I F -
b L K I L K I Q L Y M E C L R L P V T F S -
c * R S * K S S C I W S V Y D F Q * H F Q -
aagggatcagccgtgtgtatgtatagcatgagtgatgtgagaagggtgttccttggtcca
1021 ---------+---------+---------+---------+---------+---------+ 1080
ttccctagtcggcacacatacatatcgtactcactacactcttcccacaaggaaccaggt
a K G S A V C M Y S M S D V R R V F L G P -
b R D Q P C V C I A * V M * E G C S L V H -
c G I S R V Y V * H E * C E K G V P W S I -
tatgcccacagggatggacccaactatcaatgggtgccttatcaaggaagagtcccctat
1081 ---------+---------+---------+---------+---------+---------+ 1140
atacgggtgtccctacctgggttgatagttacccacggaatagttccttctcaggggata
a Y A H R D G P N Y Q W V P Y Q G R V P Y -
b M P T G M D P T I N G C L I K E E S P I -
c C P Q G W T Q L S M G A L S R K S P L S -
ccacggccaggaacttgtcccagcaaaacatttggtggttttgactctacaaaggacctt
1141 ---------+---------+---------+---------+---------+---------+ 1200
ggtgccggtccttgaacagggtcgttttgtaaaccaccaaaactgagatgtttcctggaa
a P R P G T C P S K T F G G F D S T K D L -
b H G Q E L V P A K H L V V L T L Q R T F -
c T A R N L S Q Q N I W W F * L Y K G P S -
cctgatgatgttataacctttgcaagaagtcatccagccatgtacaatccagtgtttcct
1201 ---------+---------+---------+---------+---------+---------+ 1260
ggactactacaatattggaaacgttcttcagtaggtcggtacatgttaggtcacaaagga

Figure 16-5

```
a        P  D  D  V  I  T  F  A  R  S  H  P  A  M  Y  N  P  V  F  P  -
b         L  M  M  L  *  P  L  Q  E  V  I  Q  P  C  T  I  Q  C  F  L  -
c          *  *  C  Y  N  L  C  K  K  S  S  S  H  V  Q  S  S  V  S  Y  -

         atgaacaatcgcccaatagtgatcaaaacggatgtaaattatcaatttacacaaattgtc
    1261 ---------+---------+---------+---------+---------+---------+---------+ 1320
         tacttgttagcgggttatcactagttttgcctacatttaatagttaaatgtgtttaacag

a        M  N  N  R  P  I  V  I  K  T  D  V  N  Y  Q  F  T  Q  I  V  -
b         *  T  I  A  Q  *  *  S  K  R  M  *  I  I  N  L  H  K  L  S  -
c          E  Q  S  P  N  S  D  Q  N  G  C  K  L  S  I  Y  T  N  C  R  -

         gtagaccgagtggatgcagaagatggacagtatgatgttatgtttatcggaacagatgtt
    1321 ---------+---------+---------+---------+---------+---------+---------+ 1380
         catctggctcacctacgtcttctacctgtcatactacaatacaaatagccttgtctacaa

a        V  D  R  V  D  A  E  D  G  Q  Y  D  V  M  F  I  G  T  D  V  -
b         *  T  E  W  M  Q  K  M  D  S  M  M  L  C  L  S  E  Q  M  L  -
c          R  P  S  G  C  R  R  W  T  V  *  C  Y  V  Y  R  N  R  C  W  -

         gggaccgttcttaaagtagtttcaattcctaaggagacttggtatgatttagaagaggtt
    1381 ---------+---------+---------+---------+---------+---------+---------+ 1440
         ccctggcaagaatttcatcaaagttaaggattcctctgaaccatactaaatcttctccaa

a        G  T  V  L  K  V  V  S  I  P  K  E  T  W  Y  D  L  E  E  V  -
b         G  P  F  L  K  *  F  Q  F  L  R  R  L  G  M  I  *  K  R  F  -
c          D  R  S  *  S  S  F  N  S  *  G  D  L  V  *  F  R  R  G  S  -

         ctgctggaagaaatgacagtttttcgggaaccgactgctatttcagcaatggagctttcc
    1441 ---------+---------+---------+---------+---------+---------+---------+ 1500
         gacgaccttctttactgtcaaaaagcccttggctgacgataaagtcgttacctcgaaagg

a        L  L  E  E  M  T  V  F  R  E  P  T  A  I  S  A  M  E  L  S  -
b         C  W  K  K  *  Q  F  F  G  N  R  L  L  F  Q  Q  W  S  F  P  -
c          A  G  R  N  D  S  F  S  G  T  D  C  Y  F  S  N  G  A  F  H  -

         actaagcagcaacaactatatattggttcaacggctggggttgcccagctccctttacac
    1501 ---------+---------+---------+---------+---------+---------+---------+ 1560
         tgattcgtcgttgttgatatataaccaagttgccgaccccaacgggtcgagggaaatgtg

a        T  K  Q  Q  Q  L  Y  I  G  S  T  A  G  V  A  Q  L  P  L  H  -
b         L  S  S  N  N  Y  I  L  V  Q  R  L  G  L  P  S  S  L  Y  T  -
c          *  A  A  T  T  I  Y  W  F  N  G  W  G  C  P  A  P  F  T  P  -
```

Figure 16-6

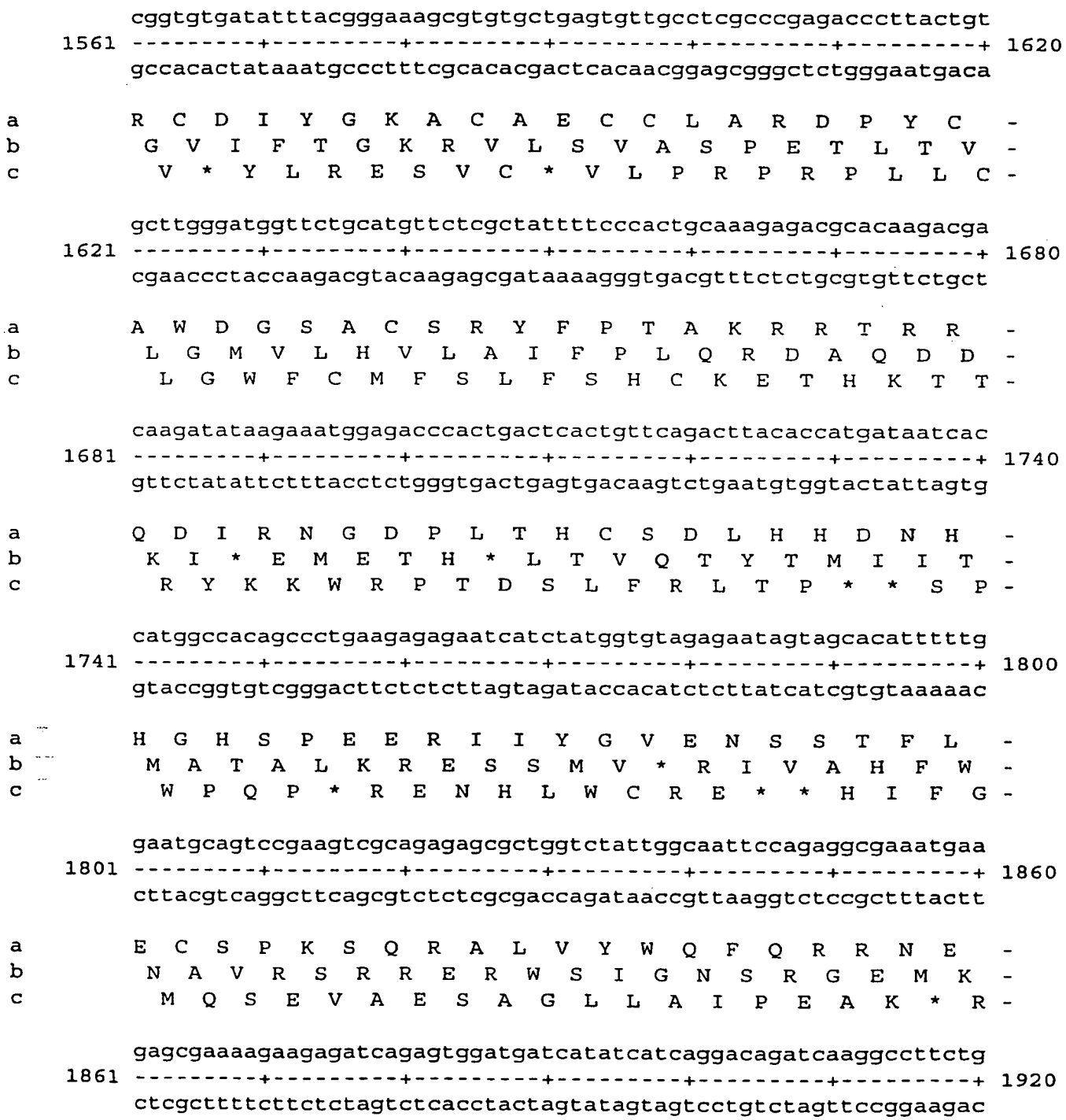
cggtgtgatatttacgggaaagcgtgtgctgagtgttgcctcgcccgagacccttactgt
1561 ---------+---------+---------+---------+---------+---------+ 1620
gccacactataaatgccctttcgcacacgactcacaacggagcgggctctgggaatgaca
a R C D I Y G K A C A E C C L A R D P Y C -
b G V I F T G K R V L S V A S P E T L T V -
c V * Y L R E S V C * V L P R P R P L L C -
gcttgggatggttctgcatgttctcgctattttcccactgcaaagagacgcacaagacga
1621 ---------+---------+---------+---------+---------+---------+ 1680
cgaaccctaccaagacgtacaagagcgataaaagggtgacgtttctctgcgtgttctgct
a A W D G S A C S R Y F P T A K R R T R R -
b L G M V L H V L A I F P L Q R D A Q D D -
c L G W F C M F S L F S H C K E T H K T T -
caagatataagaaatggagacccactgactcactgttcagacttacaccatgataatcac
1681 ---------+---------+---------+---------+---------+---------+ 1740
gttctatattctttacctctgggtgactgagtgacaagtctgaatgtggtactattagtg
a Q D I R N G D P L T H C S D L H H D N H -
b K I * E M E T H * L T V Q T Y T M I I T -
c R Y K K W R P T D S L F R L T P * * S P -
catggccacagccctgaagagagaatcatctatggtgtagagaatagtagcacattttg
1741 ---------+---------+---------+---------+---------+---------+ 1800
gtaccggtgtcgggacttctctcttagtagataccacatctcttatcatcgtgtaaaaac
a H G H S P E E R I I Y G V E N S S T F L -
b M A T A L K R E S S M V * R I V A H F W -
c W P Q P * R E N H L W C R E * * H I F G -
gaatgcagtccgaagtcgcagagagcgctggtctattggcaattccagaggcgaaatgaa
1801 ---------+---------+---------+---------+---------+---------+ 1860
cttacgtcaggcttcagcgtctctcgcgaccagataaccgttaaggtctccgctttactt
a E C S P K S Q R A L V Y W Q F Q R R N E -
b N A V R S R R E R W S I G N S R G E M K -
c M Q S E V A E S A G L L A I P E A K * R -
gagcgaaaagaagagatcagagtggatgatcatatcatcaggacagatcaaggccttctg
1861 ---------+---------+---------+---------+---------+---------+ 1920
ctcgcttttcttctctagtctcacctactagtatagtagtcctgtctagttccggaagac

Figure 16-7

```
a        E  R  K  E  E  I  R  V  D  D  H  I  I  R  T  D  Q  G  L  L  -
b         S  E  K  K  R  S  E  W  M  I  I  S  S  G  Q  I  K  A  F  C  -
c          A  K  R  R  D  Q  S  G  *  S  Y  H  Q  D  R  S  R  P  S  A  -

         ctacgtagtctacaacagaaggattcaggcaattacctctgccatgcggtggaacatggg
    1921 ---------+---------+---------+---------+---------+---------+ 1980
         gatgcatcagatgttgtcttcctaagtccgttaatggagacggtacgccaccttgtaccc

a        L  R  S  L  Q  Q  K  D  S  G  N  Y  L  C  H  A  V  E  H  G  -
b         Y  V  V  Y  N  R  R  I  Q  A  I  T  S  A  M  R  W  N  M  G  -
c          T  *  S  T  T  E  G  F  R  Q  L  P  L  P  C  G  G  T  W  V  -

         ttcatacaaactcttcttaaggtaaccctggaagtcattgacacagagcatttggaagaa
    1981 ---------+---------+---------+---------+---------+---------+ 2040
         aagtatgtttgagaagaattccattgggaccttcagtaactgtgtctcgtaaaccttctt

a        F  I  Q  T  L  L  K  V  T  L  E  V  I  D  T  E  H  L  E  E  -
b         S  Y  K  L  F  L  R  *  P  W  K  S  L  T  Q  S  I  W  K  N  -
c          H  T  N  S  S  *  G  N  P  G  S  H  *  H  R  A  F  G  R  T  -

         cttcttcataaagatgatgatggagatggctctaagaccaaagaaatgtccaatagcatg
    2041 ---------+---------+---------+---------+---------+---------+ 2100
         gaagaagtatttctactactacctctaccgagattctggtttctttacaggttatcgtac

a        L  L  H  K  D  D  D  G  D  G  S  K  T  K  E  M  S  N  S  M  -
b         F  F  I  K  M  M  M  E  M  A  L  R  P  K  K  C  P  I  A  *  -
c          S  S  *  R  *  *  W  R  W  L  *  D  Q  R  N  V  Q  *  H  D  -

         acacctagccagaaggtctggtacagagacttcatgcagctcatcaaccaccccaatctc
    2101 ---------+---------+---------+---------+---------+---------+ 2160
         tgtggatcggtcttccagaccatgtctctgaagtacgtcgagtagttggtggggttagag

a        T  P  S  Q  K  V  W  Y  R  D  F  M  Q  L  I  N  H  P  N  L  -
b         H  L  A  R  R  S  G  T  E  T  S  C  S  S  S  T  T  P  I  S  -
c          T  *  P  E  G  L  V  Q  R  L  H  A  A  H  Q  P  P  Q  S  Q  -

         aacacgatggatgagttctgtgaacaagtttggaaaagggaccgaaaacaacgtcggcaa
    2161 ---------+---------+---------+---------+---------+---------+ 2220
         ttgtgctacctactcaagacacttgttcaaacctttccctggcttttgttgcagccgtt

a        N  T  M  D  E  F  C  E  Q  V  W  K  R  D  R  K  Q  R  R  Q  -
b         T  R  W  M  S  S  V  N  K  F  G  K  G  T  E  N  N  V  G  K  -
c          H  D  G  *  V  L  *  T  S  L  E  K  G  P  K  T  T  S  A  K  -
```

Figure 16-8

```
     aggccaggacataccccagggaacagtaacaaatggaagcacttacaagaaaataagaaa
2221 ---------+---------+---------+---------+---------+---------+ 2280
     tccggtcctgtatggggtcccttgtcattgtttaccttcgtgaatgttcttttattcttt

a    R  P  G  H  T  P  G  N  S  N  K  W  K  H  L  Q  E  N  K  K  -
b     G  Q  D  I  P  Q  G  T  V  T  N  G  S  T  Y  K  K  I  R  K  -
c      A  R  T  Y  P  R  E  Q  *  Q  M  E  A  L  T  R  K  *  E  R -

     ggtagaaacaggaggacccacgaatttgagagggcacccaggagtgtctgagctgcatta
2281 ---------+---------+---------+---------+---------+---------+ 2340
     ccatctttgtcctcctgggtgcttaaactctcccgtgggtcctcacagactcgacgtaat

a    G  R  N  R  R  T  H  E  F  E  R  A  P  R  S  V  *  A  A  L  -
b     V  E  T  G  G  P  T  N  L  R  G  H  P  G  V  S  E  L  H  Y  -
c      *  K  Q  E  D  P  R  I  *  E  G  T  Q  E  C  L  S  C  I  T -

     cctctagaaacctcaaacaagtagaaacttgcctagacaataactggaaaaacaaatgca
2341 ---------+---------+---------+---------+---------+---------+ 2400
     ggagatctttggagtttgttcatctttgaacggatctgttattgacctttttgtttacgt

a    P  L  E  T  S  N  K  *  K  L  A  *  T  I  T  G  K  T  N  A  -
b     L  *  K  P  Q  T  S  R  N  L  P  R  Q  *  L  E  K  Q  M  Q  -
c      S  R  N  L  K  Q  V  E  T  C  L  D  N  N  W  K  N  K  C  N -

     atatacatgaacttttttcatggcattatgtggatgtttacaatggtgggaaattcagct
2401 ---------+---------+---------+---------+---------+---------+ 2460
     tatatgtacttgaaaaaagtaccgtaatacacctacaaatgttaccacccttaagtcga

a    I  Y  M  N  F  F  H  G  I  M  W  M  F  T  M  V  G  N  S  A  -
b     Y  T  *  T  F  F  M  A  L  C  G  C  L  Q  W  W  E  I  Q  L  -
c      I  H  E  L  F  S  W  H  Y  V  D  V  Y  N  G  G  K  F  S  * -

     gagttccaccaattataaattaaatccatgagtaactttcctaataggctttttttttcct
2461 ---------+---------+---------+---------+---------+---------+ 2520
     ctcaaggtggttaatatttaatttaggtactcattgaaaggattatccgaaaaaaaagga

a    E  F  H  Q  L  *  I  K  S  M  S  N  F  P  N  R  L  F  F  P  -
b     S  S  T  N  Y  K  L  N  P  *  V  T  F  L  I  G  F  F  F  L  -
c      V  P  P  I  I  N  *  I  H  E  *  L  S  *  *  A  F  F  S  * -

     aataccaccg
2521 ---------+ 2530
     ttatggtggc
```

Figure 16-9

```
a          N  T  T    -
b           I  P  P   -
c            Y  H     -
```

Enzymes that do cut:

NONE

Enzymes that do not cut:

NotI

Figure 17-1 HUPF-1

(Linear) MAP of: hsu59323.em_hum2 check: 6138 from: 1 to: 3602

RL;HSU59323 - Human homolog of yeast UPF1 (HUPF-I) mRNA, complete cds.
ID HSU59323 standard; RNA; HUM; 3602 BP.
XX AC U59323;
XX NI g1633577 . . .

```
        gcggcggctcggcactgttacctctcggtccggctggcgccgcgggcggtttggtccttt
      1 ---------+---------+---------+---------+---------+---------+ 60
        cgccgccgagccgtgacaatggagagccaggccgaccgcggcgcccgccaaaccaggaaa

a       A  A  A  R  H  C  Y  L  S  V  R  L  A  P  R  A  V  W  S  F  -
b        R  R  L  G  T  V  T  S  R  S  G  W  R  R  G  R  F  G  P  F  -
c         G  G  S  A  L  L  P  L  G  P  A  G  A  A  G  G  L  V  L  S -

        ccgggcgcgcgggggcgacagcggcagcgacccgaggcctgcggcctaggcctcagcgcg
     61 ---------+---------+---------+---------+---------+---------+ 120
        ggcccgcgcgcccccgctgtcgccgtcgctgggctccggacgccggatccggagtcgcgc

a       P  G  A  R  G  R  Q  R  Q  R  P  E  A  C  G  L  G  L  S  A  -
b        R  A  R  G  G  D  S  G  S  D  P  R  P  A  A  *  A  S  A  R  -
c         G  R  A  G  A  T  A  A  A  T  R  G  L  R  P  R  P  Q  R  G -

        gcggcgggctcgagtgcagcgcggaaccggcccgagggccctacccggaggcaccatgag
    121 ---------+---------+---------+---------+---------+---------+ 180
        cgccgcccgagctcacgtcgcgccttggccgggctcccgggatgggcctccgtggtactc

a       A  A  G  S  S  A  A  R  N  R  P  E  G  P  T  R  R  H  H  E  -
b        R  R  A  R  V  Q  R  G  T  G  P  R  A  L  P  G  G  T  M  S  -
c         G  G  L  E  C  S  A  E  P  A  R  G  P  Y  P  E  A  P  *  A -

        cgtggaggcgtacgggcccagctcgcagactctcactttcctggacacggaggaggccga
    181 ---------+---------+---------+---------+---------+---------+ 240
        gcacctccgcatgcccgggtcgagcgtctgagagtgaaaggacctgtgcctcctccggct

a       R  G  G  V  R  A  Q  L  A  D  S  H  F  P  G  H  G  G  G  R  -
b        V  E  A  Y  G  P  S  S  Q  T  L  T  F  L  D  T  E  E  A  E  -
c         W  R  R  T  G  P  A  R  R  L  S  L  S  W  T  R  R  R  P  S -

        gctgcttggcgccgacacacagggctccgagttcgagttcaccgactttactcttcctag
    241 ---------+---------+---------+---------+---------+---------+ 300
        cgacgaaccgcggctgtgtgtcccgaggctcaagctcaagtggctgaaatgagaaggatc
```

Figure 17-2

```
a        A  A  W  R  R  H  T  G  L  R  V  R  V  H  R  L  Y  S  S  *  -
b         L  L  G  A  D  T  Q  G  S  E  F  E  F  T  D  F  T  L  P  S -
c          C  L  A  P  T  H  R  A  P  S  S  S  S  P  T  L  L  F  L  A -

         ccagacgcagacgccccccggcggccccggcggcccgggcggtggcggcgcgggaagccc
     301 ---------+---------+---------+---------+---------+---------+ 360
         ggtctgcgtctgcgggggccgccggggccgccgggcccgccaccgccgcgcccttcggg

a        P  D  A  D  A  P  R  R  P  R  R  P  G  R  W  R  R  G  K  P  -
b         Q  T  Q  T  P  P  G  G  P  G  G  P  G  G  G  G  A  G  S  P -
c          R  R  R  R  P  P  A  A  P  A  A  R  A  V  A  A  R  E  A  R -

         gggcggcgcgggcgccggcgctgcggcgggacagctcgacgcgcaggttgggcccgaagg
     361 ---------+---------+---------+---------+---------+---------+ 420
         cccgccgcgcccgcggccgcgacgccgccctgtcgagctgcgcgtccaacccgggcttcc

a        G  R  R  G  R  R  R  C  G  G  T  A  R  R  A  G  W  A  R  R  -
b         G  G  A  G  A  G  A  A  A  G  Q  L  D  A  Q  V  G  P  E  G -
c          A  A  R  A  P  A  L  R  R  D  S  S  T  R  R  L  G  P  K  A -

         catcctgcagaacggggctgtggacgacagtgtagccaagaccagccagttgttggctga
     421 ---------+---------+---------+---------+---------+---------+ 480
         gtaggacgtcttgccccgacacctgctgtcacatcggttctggtcggtcaacaaccgact

a        H  P  A  E  R  G  C  G  R  Q  C  S  Q  D  Q  P  V  V  G  *  -
b         I  L  Q  N  G  A  V  D  D  S  V  A  K  T  S  Q  L  L  A  E -
c          S  C  R  T  G  L  W  T  T  V  *  P  R  P  A  S  C  W  L  S -

         gttgaacttcgaggaagatgaagaagacacctattacacgaaggacctccccatacacgc
     481 ---------+---------+---------+---------+---------+---------+ 540
         caacttgaagctccttctacttcttctgtggataatgtgcttcctggaggggtatgtgcg

a        V  E  L  R  G  R  *  R  R  H  L  L  H  E  G  P  P  H  T  R  -
b         L  N  F  E  E  D  E  E  D  T  Y  Y  T  K  D  L  P  I  H  A -
c          *  T  S  R  K  M  K  K  T  P  I  T  R  R  T  S  P  Y  T  P -

         ctgcagttactgtggaatacacgatcctgcctgcgtggtttactgtaataccagcaagaa
     541 ---------+---------+---------+---------+---------+---------+ 600
         gacgtcaatgacaccttatgtgctaggacggacgcaccaaatgacattatggtcgttctt

a        L  Q  L  L  W  N  T  R  S  C  L  R  G  L  L  *  Y  Q  Q  E  -
b         C  S  Y  C  G  I  H  D  P  A  C  V  V  Y  C  N  T  S  K  K -
c          A  V  T  V  E  Y  T  I  L  P  A  W  F  T  V  I  P  A  R  S -
```

Figure 17-3

```
        gtggttctgcaacggacgtggaaatacttctggcagccacattgtaaatcaccttgtgag
    601 ---------+---------+---------+---------+---------+---------+ 660
        caccaagacgttgcctgcacctttatgaagaccgtcggtgtaacatttagtggaacactc

a       V  V  L  Q  R  T  W  K  Y  F  W  Q  P  H  C  K  S  P  C  E  -
b        W  F  C  N  G  R  G  N  T  S  G  S  H  I  V  N  H  L  V  R  -
c         G  S  A  T  D  V  E  I  L  L  A  A  T  L  *  I  T  L  *  G -

        ggcaaaatgcaaagaggtgaccctgcacaaggacgggcccctgggggagacagtcctgga
    661 ---------+---------+---------+---------+---------+---------+ 720
        ccgttttacgtttctccactgggacgtgttcctgcccggggacccccctctgtcaggacct

a       G  K  M  Q  R  G  D  P  A  Q  G  R  A  P  G  G  D  S  P  G  -
b        A  K  C  K  E  V  T  L  H  K  D  G  P  L  G  E  T  V  L  E  -
c         Q  N  A  K  R  *  P  C  T  R  T  G  P  W  G  R  Q  S  W  S -

        gtgctacaactgcggctgtcgcaacgtcttcctcctcggcttcatcccggccaaagctga
    721 ---------+---------+---------+---------+---------+---------+ 780
        cacgatgttgacgccgacagcgttgcagaaggaggagccgaagtagggccggtttcgact

a       V  L  Q  L  R  L  S  Q  R  L  P  P  R  L  H  P  G  Q  S  *  -
b        C  Y  N  C  G  C  R  N  V  F  L  L  G  F  I  P  A  K  A  D  -
c         A  T  T  A  A  V  A  T  S  S  S  S  A  S  S  R  P  K  L  T -

        ctcagtggtggtgctgctgtgcaggcagccctgtgccagccagagcagcctcaaggacat
    781 ---------+---------+---------+---------+---------+---------+ 840
        gagtcaccaccacgacgacacgtccgtcgggacacggtcggtctcgtcggagttcctgta

a       L  S  G  G  A  A  V  Q  A  A  L  C  Q  P  E  Q  P  Q  G  H  -
b        S  V  V  V  L  L  C  R  Q  P  C  A  S  Q  S  S  L  K  D  I  -
c         Q  W  W  C  C  C  A  G  S  P  V  P  A  R  A  A  S  R  T  S -

        caactgggacagctcgcagtggcagccgctgatccaggaccgctgcttcctgtcctggct
    841 ---------+---------+---------+---------+---------+---------+ 900
        gttgaccctgtcgagcgtcaccgtcggcgactaggtcctggcgacgaaggacaggaccga

a       Q  L  G  Q  L  A  V  A  A  A  D  P  G  P  L  L  P  V  L  A  -
b        N  W  D  S  S  Q  W  Q  P  L  I  Q  D  R  C  F  L  S  W  L  -
c         T  G  T  A  R  S  G  S  R  *  S  R  T  A  A  S  C  P  G  W -

        ggtcaagatcccctccgagcaggagcagctgcgggcacgccagatcacggcacagcagat
    901 ---------+---------+---------+---------+---------+---------+ 960
        ccagttctaggggaggctcgtcctcgtcgacgcccgtgcggtctagtgccgtgtcgtcta
```

Figure 17-4

```
a       G  Q  D  P  L  R  A  G  A  A  A  G  T  P  D  H  G  T  A  D  -
b        V  K  I  P  S  E  Q  E  Q  L  R  A  R  Q  I  T  A  Q  Q  I  -
c         S  R  S  P  P  S  R  S  S  C  G  H  A  R  S  R  H  S  R  S  -

        caacaagctggaggagctgtggaaggaaaaccccttctgccacgctggaggacctggagaa
   961  ---------+---------+---------+---------+---------+---------+  1020
        gttgttcgacctcctcgacaccttccttttgggaagacggtgcgacctcctggacctctt

a       Q  Q  A  G  G  A  V  E  G  K  P  F  C  H  A  G  G  P  G  E  -
b        N  K  L  E  E  L  W  K  E  N  P  S  A  T  L  E  D  L  E  K  -
c         T  S  W  R  S  C  G  R  K  T  L  L  P  R  W  R  T  W  R  S  -

        gccgggggtggacgaggagccgcagcatgtcctcctgcggtacgaggacgcctaccagta
  1021  ---------+---------+---------+---------+---------+---------+  1080
        cggcccccacctgctcctcggcgtcgtacaggaggacgccatgctcctgcggatggtcat

a       A  G  G  G  R  G  A  A  A  C  P  P  A  V  R  G  R  L  P  V  -
b        P  G  V  D  E  E  P  Q  H  V  L  L  R  Y  E  D  A  Y  Q  Y  -
c         R  G  W  T  R  S  R  S  M  S  S  C  G  T  R  T  P  T  S  T  -

        ccagaacatattcgggcccctggtcaagctggaggccgactacgacaagaagctgaagga
  1081  ---------+---------+---------+---------+---------+---------+  1140
        ggtcttgtataagcccggggaccagttcgacctccggctgatgctgttcttcgacttcct

a       P  E  H  I  R  A  P  G  Q  A  G  G  R  L  R  Q  E  A  E  G  -
b        Q  N  I  F  G  P  L  V  K  L  E  A  D  Y  D  K  K  L  K  E  -
c         R  T  Y  S  G  P  W  S  S  W  R  P  T  T  T  R  S  *  R  S  -

        gtcccagactcaagataacatcactgtcaggtgggacctgggccttaacaagaagagaat
  1141  ---------+---------+---------+---------+---------+---------+  1200
        cagggtctgagttctattgtagtgacagtccaccctggacccggaattgttcttctctta

a       V  P  D  S  R  *  H  H  C  Q  V  G  P  G  P  *  Q  E  E  N  -
b        S  Q  T  Q  D  N  I  T  V  R  W  D  L  G  L  N  K  K  R  I  -
c         P  R  L  K  I  T  S  L  S  G  G  T  W  A  L  T  R  R  E  S  -

        cgcctacttcactttgcccaagactgactctgacatgcggctcatgcaggggatgagat
  1201  ---------+---------+---------+---------+---------+---------+  1260
        gcggatgaagtgaaacgggttctgactgagactgtacgccgagtacgtccccctactcta

a       R  L  L  H  F  A  Q  D  *  L  *  H  A  A  H  A  G  G  *  D  -
b        A  Y  F  T  L  P  K  T  D  S  D  M  R  L  M  Q  G  D  E  I  -
c         P  T  S  L  C  P  R  L  T  L  T  C  G  S  C  R  G  M  R  Y  -
```

Figure 17-5

```
          atgcctgcggtacaaaggggaccttgcgcccctgtggaaagggatcggccacgtcatcaa
     1261 ---------+---------+---------+---------+---------+---------+ 1320
          tacggacgccatgtttcccctggaacgcggggacacctttccctagccggtgcagtagtt

a         M  P  A  V  Q  R  G  P  C  A  P  V  E  R  D  R  P  R  H  Q   -
b          C  L  R  Y  K  G  D  L  A  P  L  W  K  G  I  G  H  V  I  K  -
c           A  C  G  T  K  G  T  L  R  P  C  G  K  G  S  A  T  S  S  R -

          ggtccctgataattatggcgatgagatcgccattgagctgcggagcagcgtgggtgcacc
     1321 ---------+---------+---------+---------+---------+---------+ 1380
          ccagggactattaataccgctactctagcggtaactcgacgcctcgtcgcacccacgtgg

a         G  P  *  *  L  W  R  *  D  R  H  *  A  A  E  Q  R  G  C  T   -
b          V  P  D  N  Y  G  D  E  I  A  I  E  L  R  S  S  V  G  A  P  -
c           S  L  I  I  M  A  M  R  S  P  L  S  C  G  A  A  W  V  H  L -

          tgtggaggtgactcacaacttccaggtggattttgtgtggaagtcgacctcctttgacag
     1381 ---------+---------+---------+---------+---------+---------+ 1440
          acacctccactgagtgttgaaggtccacctaaaacacaccttcagctggaggaaactgtc

a         C  G  G  D  S  Q  L  P  G  G  F  C  V  E  V  D  L  L  *  Q   -
b          V  E  V  T  H  N  F  Q  V  D  F  V  W  K  S  T  S  F  D  R  -
c           W  R  *  L  T  T  S  R  W  I  L  C  G  S  R  P  P  L  T  G -

          gatgcagagcgcattgaaaacgtttgccgtggatgagacctcggtgtctggctacatcta
     1441 ---------+---------+---------+---------+---------+---------+ 1500
          ctacgtctcgcgtaacttttgcaaacggcacctactctggagccacagaccgatgtagat

a         D  A  E  R  I  E  N  V  C  R  G  *  D  L  G  V  W  L  H  L   -
b          M  Q  S  A  L  K  T  F  A  V  D  E  T  S  V  S  G  Y  I  Y  -
c           C  R  A  H  *  K  R  L  P  W  M  R  P  R  C  L  A  T  S  T -

          ccacaagctgttgggccacgaggtggaggacgtaatcaccaagtgccagctgcccaagcg
     1501 ---------+---------+---------+---------+---------+---------+ 1560
          ggtgttcgacaacccggtgctccacctcctgcattagtggttcacggtcgacgggttcgc

a         P  Q  A  V  G  P  R  G  G  G  R  N  H  Q  V  P  A  A  Q  A   -
b          H  K  L  L  G  H  E  V  E  D  V  I  T  K  C  Q  L  P  K  R  -
c           T  S  C  W  A  T  R  W  R  T  *  S  P  S  A  S  C  P  S  A -

          cttcacggcgcagggcctccccgacctcaaccactcccaggtttatgccgtgaagactgt
     1561 ---------+---------+---------+---------+---------+---------+ 1620
          gaagtgccgcgtcccggaggggctggagttggtgagggtccaaatacggcacttctgaca
```

Figure 17-6

```
a       L  H  G  A  G  P  P  R  P  Q  P  L  P  G  L  C  R  E  D  C  -
b        F  T  A  Q  G  L  P  D  L  N  H  S  Q  V  Y  A  V  K  T  V  -
c         S  R  R  R  A  S  P  T  S  T  T  P  R  F  M  P  *  R  L  C  -

         gctgcaaagaccactgagcctgatccagggcccgccaggcacggggaagacggtgacgtc
    1621 ---------+---------+---------+---------+---------+---------+ 1680
         cgacgtttctggtgactcggactaggtcccgggcggtccgtgccccttctgccactgcag

a       A  A  K  T  T  E  P  D  P  G  P  A  R  H  G  E  D  G  D  V  -
b        L  Q  R  P  L  S  L  I  Q  G  P  P  G  T  G  K  T  V  T  S  -
c         C  K  D  H  *  A  *  S  R  A  R  Q  A  R  G  R  R  *  R  R  -

         ggccaccatcgtctaccacctggcccggcaaggcaacgggccggtgctggtgtgtgctcc
    1681 ---------+---------+---------+---------+---------+---------+ 1740
         ccggtggtagcagatggtggaccgggccgttccgttgcccggccacgaccacacacgagg

a       G  H  H  R  L  P  P  G  P  A  R  Q  R  A  G  A  G  V  C  S  -
b        A  T  I  V  Y  H  L  A  R  Q  G  N  G  P  V  L  V  C  A  P  -
c         P  P  S  S  T  T  W  P  G  K  A  T  G  R  C  W  C  V  L  R  -

         gagcaacatcgccgtggaccagctaacggagaagatccaccagacggggctaaaggtcgt
    1741 ---------+---------+---------+---------+---------+---------+ 1800
         ctcgttgtagcggcacctggtcgattgcctcttctaggtggtctgccccgatttccagca

a       E  Q  H  R  R  G  P  A  N  G  E  D  P  P  D  G  A  K  G  R  -
b        S  N  I  A  V  D  Q  L  T  E  K  I  H  Q  T  G  L  K  V  V  -
c         A  T  S  P  W  T  S  *  R  R  R  S  T  R  R  G  *  R  S  C  -

         gcgcctctgcgccaagagccgtgaggccatcgactccccggtgtcttttctggccctgca
    1801 ---------+---------+---------+---------+---------+---------+ 1860
         cgcggagacgcggttctcggcactccggtagctgaggggccacagaaaagaccgggacgt

a       A  P  L  R  Q  E  P  *  G  H  R  L  P  G  V  F  S  G  P  A  -
b        R  L  C  A  K  S  R  E  A  I  D  S  P  V  S  F  L  A  L  H  -
c         A  S  A  P  R  A  V  R  P  S  T  P  R  C  L  F  W  P  C  T  -

         caaccagatcaggaacatggacagcatgcctgagctgcagaagctgcagcagctgaaaga
    1861 ---------+---------+---------+---------+---------+---------+ 1920
         gttggtctagtccttgtacctgtcgtacggactcgacgtcttcgacgtcgtcgactttct

a       Q  P  D  Q  E  H  G  Q  H  A  *  A  A  E  A  A  A  A  E  R  -
b        N  Q  I  R  N  M  D  S  M  P  E  L  Q  K  L  Q  Q  L  K  D  -
c         T  R  S  G  T  W  T  A  C  L  S  C  R  S  C  S  S  *  K  T  -
```

Figure 17-7

```
          cgagactggggagctgtcgtctgccgacgagaagcggtaccgggccttgaagcgcaccgc
     1921 ---------+---------+---------+---------+---------+---------+ 1980
          gctctgacccctcgacagcagacggctgctcttcgccatggcccggaacttcgcgtggcg

a         R  D  W  G  A  V  V  C  R  R  E  A  V  P  G  L  E  A  H  R  -
b          E  T  G  E  L  S  S  A  D  E  K  R  Y  R  A  L  K  R  T  A -
c           R  L  G  S  C  R  L  P  T  R  S  G  T  G  P  *  S  A  P  Q -

          agagagagagctgctgatgaacgcagatgtcatctgctgcacatgtgtgggcgccggtga
     1981 ---------+---------+---------+---------+---------+---------+ 2040
          tctctctctcgacgactacttgcgtctacagtagacgacgtgtacacacccgcggccact

a         R  E  R  A  A  D  E  R  R  C  H  L  L  H  M  C  G  R  R  *  -
b          E  R  E  L  L  M  N  A  D  V  I  C  C  T  C  V  G  A  G  D -
c           R  E  S  C  *  *  T  Q  M  S  S  A  A  H  V  W  A  P  V  T -

          cccgaggctggccaagatgcagttccgctccattttaatcgacgaaagcacccaggccac
     2041 ---------+---------+---------+---------+---------+---------+ 2100
          gggctccgaccggttctacgtcaaggcgaggtaaaattagctgctttcgtgggtccggtg

a         P  E  A  G  Q  D  A  V  P  L  H  F  N  R  R  K  H  P  G  H  -
b          P  R  L  A  K  M  Q  F  R  S  I  L  I  D  E  S  T  Q  A  T -
c           R  G  W  P  R  C  S  S  A  P  F  *  S  T  K  A  P  R  P  P -

          cgagccggagtgcatggttcccgtggtcctcggggccaagcagctgatccttgtaggcga
     2101 ---------+---------+---------+---------+---------+---------+ 2160
          gctcggcctcacgtaccaagggcaccaggagccccggttcgtcgactaggaacatccgct

a         R  A  G  V  H  G  S  R  G  P  R  G  Q  A  A  D  P  C  R  R  -
b          E  P  E  C  M  V  P  V  V  L  G  A  K  Q  L  I  L  V  G  D -
c           S  R  S  A  W  F  P  W  S  S  G  P  S  S  *  S  L  *  A  T -

          ccactgccagctgggcccagtggtgatgtgcaagaaggcggccaaggccgggctgtcaca
     2161 ---------+---------+---------+---------+---------+---------+ 2220
          ggtgacggtcgacccgggtcaccactacacgttcttccgccggttccggcccgacagtgt

a         P  L  P  A  G  P  S  G  D  V  Q  E  G  G  Q  G  R  A  V  T  -
b          H  C  Q  L  G  P  V  V  M  C  K  K  A  A  K  A  G  L  S  Q -
c           T  A  S  W  A  Q  W  *  C  A  R  R  R  P  R  P  G  C  H  S -

          gtcgctcttcgagcgcctggtggtgctgggcatccggcccatccgcctgcaggtccagta
     2221 ---------+---------+---------+---------+---------+---------+ 2280
          cagcgagaagctcgcggaccaccacgacccgtaggccgggtaggcggacgtccaggtcat
```

Figure 17-8

```
a        V  A  L  R  A  P  G  G  A  G  H  P  A  H  P  P  A  G  P  V  -
b         S  L  F  E  R  L  V  V  L  G  I  R  P  I  R  L  Q  V  Q  Y  -
c          R  S  S  S  A  W  W  C  W  A  S  G  P  S  A  C  R  S  S  T -

         ccggatgcaccctgcactcagcgccttcccatccaacatcttctacgagggctccctcca
    2281 ---------+---------+---------+---------+---------+---------+ 2340
         ggcctacgtgggacgtgagtcgcggaagggtaggttgtagaagatgctcccgagggaggt

a        P  D  A  P  C  T  Q  R  L  P  I  Q  H  L  L  R  G  L  P  P  -
b         R  M  H  P  A  L  S  A  F  P  S  N  I  F  Y  E  G  S  L  Q  -
c          G  C  T  L  H  S  A  P  S  H  P  T  S  S  T  R  A  P  S  R -

         gaatggtgtcactgcagcggatcgtgtgaagaagggatttgacttccagtggccccaacc
    2341 ---------+---------+---------+---------+---------+---------+ 2400
         cttaccacagtgacgtcgcctagcacacttcttccctaaactgaaggtcaccggggttgg

a        E  W  C  H  C  S  G  S  C  E  E  G  I  *  L  P  V  A  P  T  -
b         N  G  V  T  A  A  D  R  V  K  K  G  F  D  F  Q  W  P  Q  P  -
c          M  V  S  L  Q  R  I  V  *  R  R  D  L  T  S  S  G  P  N  P -

         cgataaaccgatgttcttctacgtgacccagggccaagaggagattgccagctcgggcac
    2401 ---------+---------+---------+---------+---------+---------+ 2460
         gctatttggctacaagaagatgcactgggtcccggttctcctctaacggtcgagcccgtg

a        R  *  T  D  V  L  L  R  D  P  G  P  R  G  D  C  Q  L  G  H  -
b         D  K  P  M  F  F  Y  V  T  Q  G  Q  E  E  I  A  S  S  G  T  -
c          I  N  R  C  S  S  T  *  P  R  A  K  R  R  L  P  A  R  A  P -

         ctcctacctgaacaggaccgaggctgcgaacgtggagaagatcaccacgaagttgctgaa
    2461 ---------+---------+---------+---------+---------+---------+ 2520
         gaggatggacttgtcctggctccgacgcttgcacctcttctagtggtgcttcaacgactt

a        L  L  P  E  Q  D  R  G  C  E  R  G  E  D  H  H  E  V  A  E  -
b         S  Y  L  N  R  T  E  A  A  N  V  E  K  I  T  T  K  L  L  K  -
c          P  T  *  T  G  P  R  L  R  T  W  R  R  S  P  R  S  C  *  R -

         ggcaggcgccaagccggaccagattggcatcatcacgccctacgagggccagcgctccta
    2521 ---------+---------+---------+---------+---------+---------+ 2580
         ccgtccgcggttcggcctggtctaaccgtagtagtgcgggatgctcccggtcgcgaggat

a        G  R  R  Q  A  G  P  D  W  H  H  H  A  L  R  G  P  A  L  L  -
b         A  G  A  K  P  D  Q  I  G  I  I  T  P  Y  E  G  Q  R  S  Y  -
c          Q  A  P  S  R  T  R  L  A  S  S  R  P  T  R  A  S  A  P  T -
```

Figure 17-9

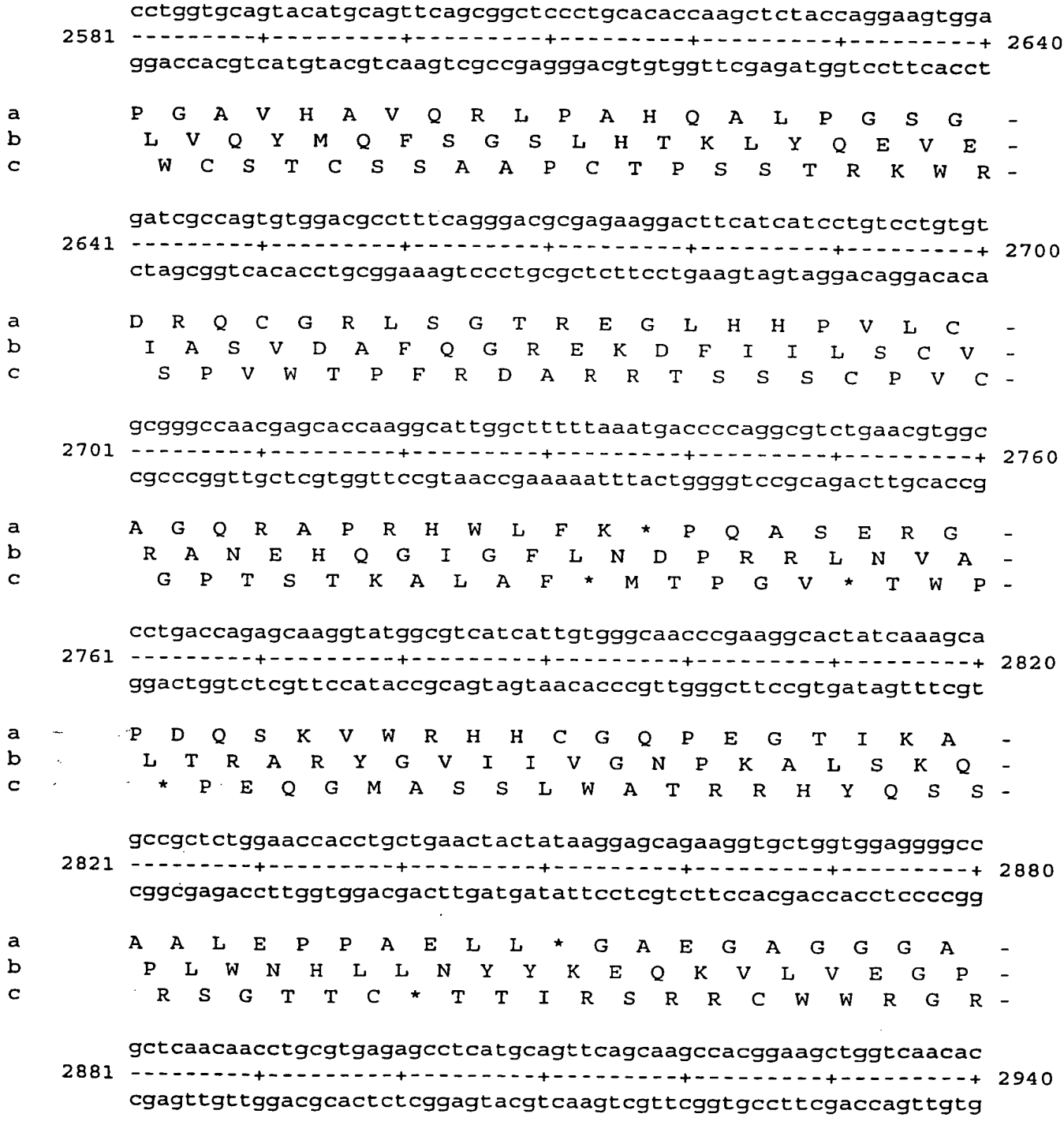
cctggtgcagtacatgcagttcagcggctccctgcacaccaagctctaccaggaagtgga
2581 ---------+---------+---------+---------+---------+---------+ 2640
ggaccacgtcatgtacgtcaagtcgccgagggacgtgtggttcgagatggtccttcacct
a P G A V H A V Q R L P A H Q A L P G S G -
b L V Q Y M Q F S G S L H T K L Y Q E V E -
c W C S T C S S A A P C T P S S T R K W R -
gatcgccagtgtggacgcctttcagggacgcgagaaggacttcatcatcctgtcctgtgt
2641 ---------+---------+---------+---------+---------+---------+ 2700
ctagcggtcacacctgcggaaagtccctgcgctcttcctgaagtagtaggacaggacaca
a D R Q C G R L S G T R E G L H H P V L C -
b I A S V D A F Q G R E K D F I I L S C V -
c S P V W T P F R D A R R T S S S C P V C -
gcgggccaacgagcaccaaggcattggcttttaaatgaccccaggcgtctgaacgtggc
2701 ---------+---------+---------+---------+---------+---------+ 2760
cgcccggttgctcgtggttccgtaaccgaaaaatttactggggtccgcagacttgcaccg
a A G Q R A P R H W L F K * P Q A S E R G -
b R A N E H Q G I G F L N D P R R L N V A -
c G P T S T K A L A F * M T P G V * T W P -
cctgaccagagcaaggtatggcgtcatcattgtgggcaacccgaaggcactatcaaagca
2761 ---------+---------+---------+---------+---------+---------+ 2820
ggactggtctcgttccataccgcagtagtaacacccgttgggcttccgtgatagtttcgt
a P D Q S K V W R H H C G Q P E G T I K A -
b L T R A R Y G V I I V G N P K A L S K Q -
c * P E Q G M A S S L W A T R R H Y Q S S -
gccgctctggaaccacctgctgaactactataaggagcagaaggtgctggtggaggggcc
2821 ---------+---------+---------+---------+---------+---------+ 2880
cggcgagaccttggtggacgacttgatgatattcctcgtcttccacgaccacctccccgg
a A A L E P P A E L L * G A E G A G G G A -
b P L W N H L L N Y Y K E Q K V L V E G P -
c R S G T T C * T T I R S R R C W W R G R -
gctcaacaacctgcgtgagagcctcatgcagttcagcaagccacggaagctggtcaacac
2881 ---------+---------+---------+---------+---------+---------+ 2940
cgagttgttggacgcactctcggagtacgtcaagtcgttcggtgccttcgaccagttgtg

Figure 17-10

```
a        A  Q  Q  P  A  *  E  P  H  A  V  Q  Q  A  T  E  A  G  Q  H  -
b         L  N  N  L  R  E  S  L  M  Q  F  S  K  P  R  K  L  V  N  T  -
c          S  T  T  C  V  R  A  S  C  S  S  A  S  H  G  S  W  S  T  L -

         tatcaacccgggagcccgcttcatgaccacagccatgtatgatgcccgggaggccatcat
    2941 ---------+---------+---------+---------+---------+---------+---------+ 3000
         atagttgggccctcgggcgaagtactggtgtcggtacatactacgggccctccggtagta

a        Y  Q  P  G  S  P  L  H  D  H  S  H  V  *  C  P  G  G  H  H  -
b         I  N  P  G  A  R  F  M  T  T  A  M  Y  D  A  R  E  A  I  I  -
c          S  T  R  E  P  A  S  *  P  Q  P  C  M  M  P  G  R  P  S  S -

         cccaggctccgtctatgatcggagcagccagggccggccttccagcatgtacttccagac
    3001 ---------+---------+---------+---------+---------+---------+---------+ 3060
         gggtccgaggcagatactagcctcgtcggtcccggccggaaggtcgtacatgaaggtctg

a        P  R  L  R  L  *  S  E  Q  P  G  P  A  F  Q  H  V  L  P  D  -
b         P  G  S  V  Y  D  R  S  S  Q  G  R  P  S  S  M  Y  F  Q  T  -
c          Q  A  P  S  M  I  G  A  A  R  A  G  L  P  A  C  T  S  R  P -

         ccatgaccagattggcatgatcagtgccggccctagccacgtggctgccatgaacattcc
    3061 ---------+---------+---------+---------+---------+---------+---------+ 3120
         ggtactggtctaaccgtactagtcacggccgggatcggtgcaccgacggtacttgtaagg

a        P  *  P  D  W  H  D  Q  C  R  P  *  P  R  G  C  H  E  H  S  -
b         H  D  Q  I  G  M  I  S  A  G  P  S  H  V  A  A  M  N  I  P  -
c          M  T  R  L  A  *  S  V  P  A  L  A  T  W  L  P  *  T  F  P -

         catccccttcaacctggtcatgccacccatgccaccgcctggctattttggacaagccaa
    3121 ---------+---------+---------+---------+---------+---------+---------+ 3180
         gtaggggaagttggaccagtacggtgggtacggtggcggaccgataaaacctgttcggtt

a        H  P  L  Q  P  G  H  A  T  H  A  T  A  W  L  F  W  T  S  Q  -
b         I  P  F  N  L  V  M  P  P  M  P  P  P  G  Y  F  G  Q  A  N  -
c          S  P  S  T  W  S  C  H  P  C  H  R  L  A  I  L  D  K  P  T -

         cgggcctgctgcagggcgaggcaccccgaaaggcaagactggtcgtgggggacgccagaa
    3181 ---------+---------+---------+---------+---------+---------+---------+ 3240
         gcccggacgacgtcccgctccgtggggctttccgttctgaccagcacccccctgcggtctt

a        R  A  C  C  R  A  R  H  P  E  R  Q  D  W  S  W  G  T  P  E  -
b         G  P  A  A  G  R  G  T  P  K  G  K  T  G  R  G  G  R  Q  K  -
c          G  L  L  Q  G  E  A  P  R  K  A  R  L  V  V  G  D  A  R  R -
```

Figure 17-11

```
          gaaccgctttgggcttcctggacccagccagactaacctccccaacagccaagccagcca
     3241 ---------+---------+---------+---------+---------+---------+ 3300
          cttggcgaaacccgaaggacctgggtcggtctgattggaggggttgtcggttcggtcggt

a         E  P  L  W  A  S  W  T  Q  P  D  *  P  P  Q  Q  P  S  Q  P  -
b          N  R  F  G  L  P  G  P  S  Q  T  N  L  P  N  S  Q  A  S  Q  -
c           T  A  L  G  F  L  D  P  A  R  L  T  S  P  T  A  K  P  A  R -

          ggatgtggcgtcacagcccttctctcagggcgccctgacgcagggctacatctccatgag
     3301 ---------+---------+---------+---------+---------+---------+ 3360
          cctacaccgcagtgtcgggaagagagtcccgcgggactgcgtcccgatgtagaggtactc

a         G  C  G  V  T  A  L  L  S  G  R  P  D  A  G  L  H  L  H  E  -
b          D  V  A  S  Q  P  F  S  Q  G  A  L  T  Q  G  Y  I  S  M  S  -
c           M  W  R  H  S  P  S  L  R  A  P  *  R  R  A  T  S  P  *  A -

          ccagccttcccagatgagccagcccggcctctcccagccggagctgtcccaggacagtta
     3361 ---------+---------+---------+---------+---------+---------+ 3420
          ggtcggaagggtctactcggtcgggccggagagggtcggcctcgacagggtcctgtcaat

a         P  A  F  P  D  E  P  A  R  P  L  P  A  G  A  V  P  G  Q  L  -
b          Q  P  S  Q  M  S  Q  P  G  L  S  Q  P  E  L  S  Q  D  S  Y  -
c           S  L  P  R  *  A  S  P  A  S  P  S  R  S  C  P  R  T  V  T -

          ccttggtgacgagtttaaatcacaaatcgacgtggcgctctcacaggactccacgtacca
     3421 ---------+---------+---------+---------+---------+---------+ 3480
          ggaaccactgctcaaatttagtgtttagctgcaccgcgagagtgtcctgaggtgcatggt

a         P  W  *  R  V  *  I  T  N  R  R  G  A  L  T  G  L  H  V  P  -
b          L  G  D  E  F  K  S  Q  I  D  V  A  L  S  Q  D  S  T  Y  Q  -
c           L  V  T  S  L  N  H  K  S  T  W  R  S  H  R  T  P  R  T  R -

          gggagagcgggcttaccagcatggcggggtgacggggctgtcccagtattaaaaggtggc
     3481 ---------+---------+---------+---------+---------+---------+ 3540
          ccctctcgcccgaatggtcgtaccgccccactgccccgacagggtcataattttccaccg

a         G  R  A  G  L  P  A  W  R  G  D  G  A  V  P  V  L  K  G  G  -
b          G  E  R  A  Y  Q  H  G  G  V  T  G  L  S  Q  Y  *  K  V  A  -
c           E  S  G  L  T  S  M  A  G  *  R  G  C  P  S  I  K  R  W  R -

          ggcggaagagctaagcaacgtggcttagtccatcagcatcttattctgggtaataaaaaa
     3541 ---------+---------+---------+---------+---------+---------+ 3600
          ccgccttctcgattcgttgcaccgaatcaggtagtcgtagaataagacccattatttttt
```

Figure 17-12

```
a        G  G  R  A  K  Q  R  G  L  V  H  Q  H  L  I  L  G  N  K  K  -
b         A  E  E  L  S  N  V  A  *  S  I  S  I  L  F  W  V  I  K  N  -
c          R  K  S  *  A  T  W  L  S  P  S  A  S  Y  S  G  *  *  K  M  -

         tg
    3601 -- 3602
         ac

a           -
b           -
c           -
```

Figure 18-1 HMG

(Linear) MAP of: hshmgicr.em_hum1 check: 9603 from: 1 to: 1200

RL;HSHMGICR - H.sapiens HMGI-C mRNA for high mobility group protein I-C
ID HSHMGICR standard; RNA; HUM; 1200 BP.
XX
AC Z31595;
XX
NI g468705 . . .

March 20, 1998 10:31 ..

```
          tcttgaatcttggggcaggaactcagaaaacttccagcccgggcagcgcgcgcttggtgc
        1 ---------+---------+---------+---------+---------+---------+ 60
          agaacttagaaccccgtccttgagtcttttgaaggtcgggcccgtcgcgcgcgaaccacg

a         S  *  I  L  G  Q  E  L  R  K  L  P  A  R  A  A  R  A  W  C  -
b          L  E  S  W  G  R  N  S  E  N  F  Q  P  G  Q  R  A  L  G  A  -
c           L  N  L  G  A  G  T  Q  K  T  S  S  P  G  S  A  R  L  V  Q -

          aagactcaggagctagcagcccgtccccctccgactctccggtgccgccgctgcctgctc
       61 ---------+---------+---------+---------+---------+---------+ 120
          ttctgagtcctcgatcgtcgggcaggggggaggctgagaggccacggcggcgacggacgag

a         K  T  Q  E  L  A  A  R  P  P  P  T  L  R  C  R  R  C  L  L  -
b          R  L  R  S  *  Q  P  V  P  L  R  L  S  G  A  A  A  A  C  S  -
c           D  S  G  A  S  S  P  S  P  S  D  S  P  V  P  P  L  P  A  P -

          ccgccaccctaggaggcgcggtgccacccactactctgtcctctgcctgtgctccgtgcc
      121 ---------+---------+---------+---------+---------+---------+ 180
          ggcggtgggatcctccgcgccacggtgggtgatgagacaggagacggacacgaggcacgg

a         P  P  P  *  E  A  R  C  H  P  L  L  C  P  L  P  V  L  R  A  -
b          R  H  P  R  R  R  G  A  T  H  Y  S  V  L  C  L  C  S  V  P  -
c           A  T  L  G  G  A  V  P  P  T  T  L  S  S  A  C  A  P  C  P -

          cgaccctatcccggcggagtctccccatcctcctttgctttccgactgcccaaggcactt
      181 ---------+---------+---------+---------+---------+---------+ 240
          gctgggatagggccgcctcagaggggtaggaggaaacgaaaggctgacgggttccgtgaa

a         R  P  Y  P  G  G  V  S  P  S  S  F  A  F  R  L  P  K  A  L  -
b          D  P  I  P  A  E  S  P  H  P  P  L  L  S  D  C  P  R  H  F  -
c           T  L  S  R  R  S  L  P  I  L  L  C  F  P  T  A  Q  G  T  F -

          tcaatctcaatctcttctctctctctctctctctctgtctctctctctctctctctct
      241 ---------+---------+---------+---------+---------+---------+ 300
          agttagagttagagaagagagagagagagagagagacagagagagagagagagagaga
```

Figure 18-2

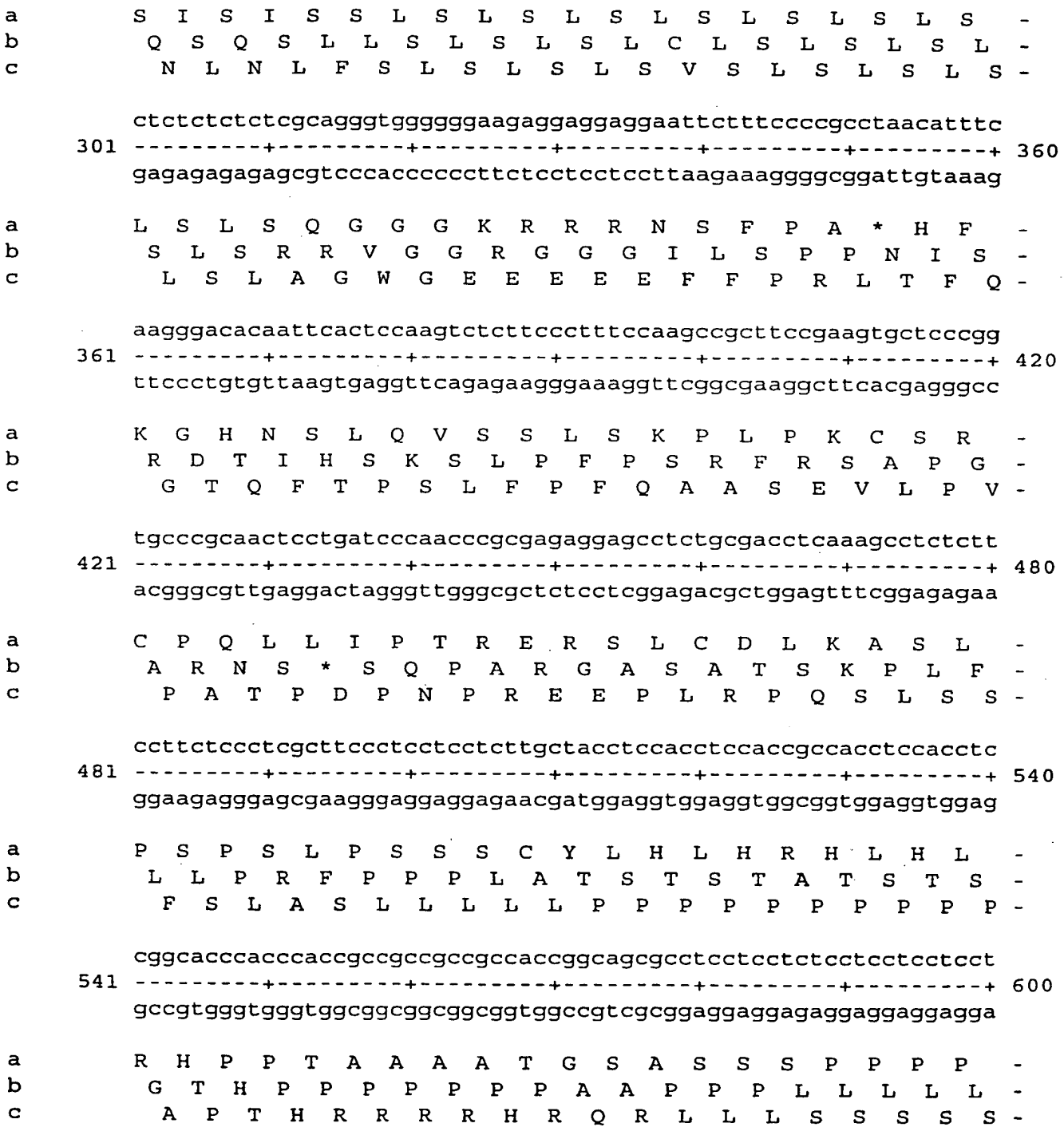
a S I S I S S L S L S L S L S L S L S L S -
b Q S Q S L L S L S L S L C L S L S L S L -
c N L N L F S L S L S L S V S L S L S L S -
ctctctctctcgcagggtgggggaagaggaggaggaattctttccccgcctaacatttc
301 ---------+---------+---------+---------+---------+---------+---------+ 360
gagagagagagcgtcccacccccttctcctcctccttaagaaaggggcggattgtaaag
a L S L S Q G G G K R R R N S F P A * H F -
b S L S R R V G G R G G G I L S P P N I S -
c L S L A G W G E E E E E F F P R L T F Q -
aagggacacaattcactccaagtctcttccctttccaagccgcttccgaagtgctcccgg
361 ---------+---------+---------+---------+---------+---------+---------+ 420
ttccctgtgttaagtgaggttcagagaagggaaaggttcggcgaaggcttcacgagggcc
a K G H N S L Q V S S L S K P L P K C S R -
b R D T I H S K S L P F P S R F R S A P G -
c G T Q F T P S L F P F Q A A S E V L P V -
tgcccgcaactcctgatcccaacccgcgagaggagcctctgcgacctcaaagcctctctt
421 ---------+---------+---------+---------+---------+---------+---------+ 480
acgggcgttgaggactagggttgggcgctctcctcggagacgctggagtttcggagagaa
a C P Q L L I P T R E R S L C D L K A S L -
b A R N S * S Q P A R G A S A T S K P L F -
c P A T P D P N P R E E P L R P Q S L S S -
ccttctccctcgcttccctcctcctcttgctacctccacctccaccgccacctccacctc
481 ---------+---------+---------+---------+---------+---------+---------+ 540
ggaagagggagcgaagggaggaggagaacgatggaggtggaggtggcggtggaggtggag
a P S P S L P S S S C Y L H L H R H L H L -
b L L P R F P P P L A T S T S T A T S T S -
c F S L A S L L L L L P P P P P P P P P P -
cggcacccacccaccgccgccgccgccaccggcagcgcctcctcctctcctcctcctcct
541 ---------+---------+---------+---------+---------+---------+---------+ 600
gccgtgggtgggtggcggcggcggcggtggccgtcgcggaggaggagaggaggaggagga
a R H P P T A A A A T G S A S S S P P P P -
b G T H P P P P P P P P A A P P P L L L L L -
c A P T H R R R R R H R Q R L L L S S S S S -

Figure 18-3

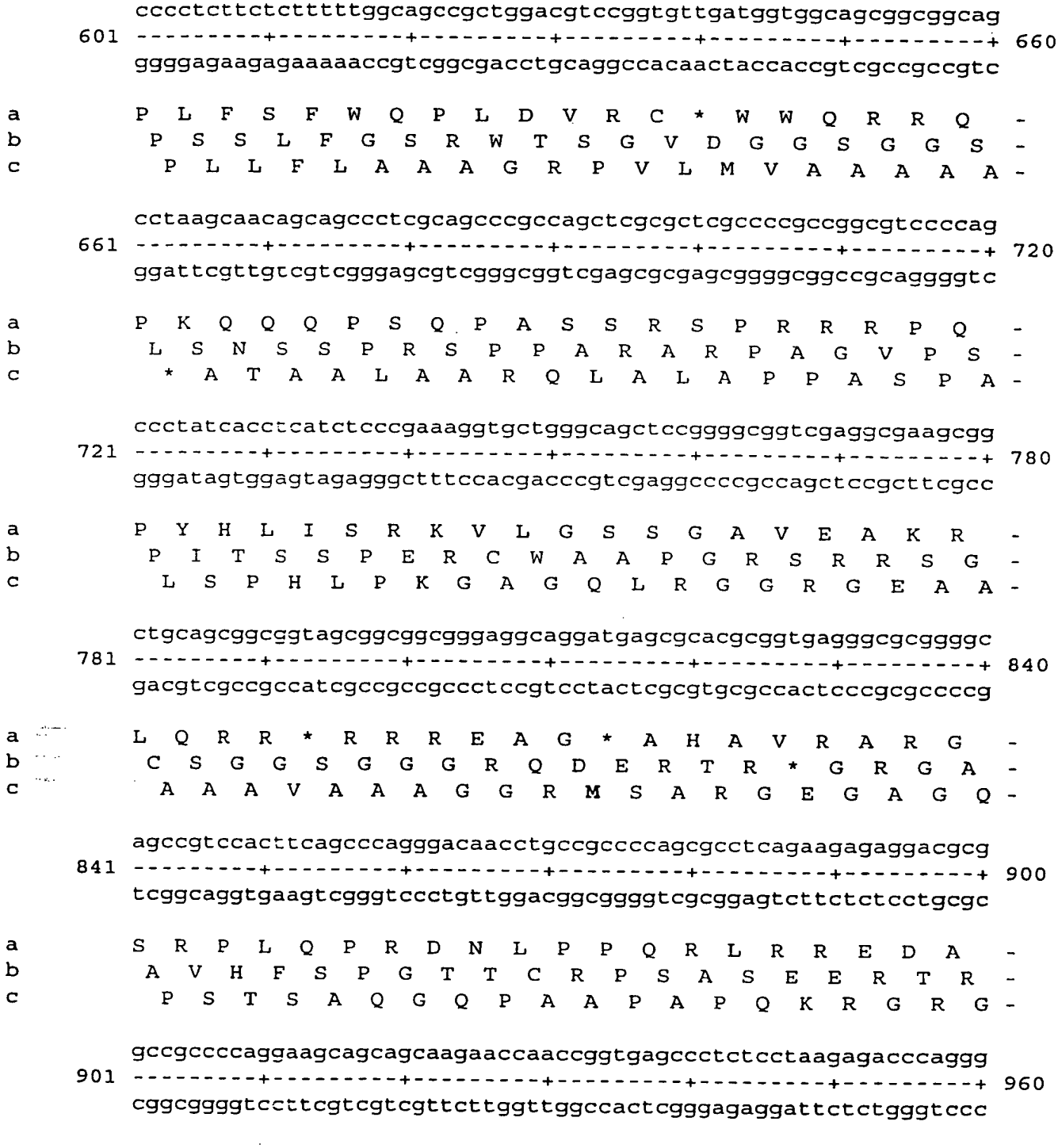
ccccctcttctcttttttggcagccgctggacgtccggtgttgatggtggcagcggcggcag
601 ---------+---------+---------+---------+---------+---------+---------+ 660
ggggagaagagaaaaaccgtcggcgacctgcaggccacaactaccaccgtcgccgccgtc
a P L F S F W Q P L D V R C * W W Q R R Q -
b P S S L F G S R W T S G V D G G S G G S -
c P L L F L A A A G R P V L M V A A A A A -
cctaagcaacagcagccctcgcagcccgccagctcgcgctcgccccgccggcgtccccag
661 ---------+---------+---------+---------+---------+---------+---------+ 720
ggattcgttgtcgtcgggagcgtcgggcggtcgagcgcgagcggggcggccgcaggggtc
a P K Q Q Q P S Q P A S S R S P R R R P Q -
b L S N S S P R S P P A R A R P A G V P S -
c * A T A A L A A R Q L A L A P P A S P A -
ccctatcacctcatctcccgaaaggtgctgggcagctccggggcggtcgaggcgaagcgg
721 ---------+---------+---------+---------+---------+---------+---------+ 780
gggatagtggagtagagggctttccacgacccgtcgaggccccgccagctccgcttcgcc
a P Y H L I S R K V L G S S G A V E A K R -
b P I T S S P E R C W A A P G R S R R S G -
c L S P H L P K G A G Q L R G G R G E A A -
ctgcagcggcggtagcggcggcgggaggcaggatgagcgcacgcggtgagggcgcggggc
781 ---------+---------+---------+---------+---------+---------+---------+ 840
gacgtcgccgccatcgccgccgccctccgtcctactcgcgtgcgccactcccgcgccccg
a L Q R R * R R R E A G * A H A V R A R G -
b C S G G S G G G R Q D E R T R * G R G A -
c A A A V A A A G G R M S A R G E G A G Q -
agccgtccacttcagcccagggacaacctgccgccccagcgcctcagaagagaggacgcg
841 ---------+---------+---------+---------+---------+---------+---------+ 900
tcggcaggtgaagtcgggtccctgttggacggcggggtcgcggagtcttctctcctgcgc
a S R P L Q P R D N L P P Q R L R R E D A -
b A V H F S P G T T C R P S A S E E R T R -
c P S T S A Q G Q P A A P A P Q K R G R G -
gccgccccaggaagcagcagcaagaaccaaccggtgagccctctcctaagagacccaggg
901 ---------+---------+---------+---------+---------+---------+---------+ 960
cggcggggtccttcgtcgtcgttcttggttggccactcgggagaggattctctgggtccc

Figure 18-4

```
a        A  A  P  G  S  S  S  K  N  Q  P  V  S  P  L  L  R  D  P  G  -
b         P  P  Q  E  A  A  A  R  T  N  R  *  A  L  S  *  E  T  Q  G  -
c          R  P  R  K  Q  Q  Q  E  P  T  G  E  P  S  P  K  R  P  R  G -

         gaagacccaaaggcagcaaaaacaagagtccctctaaagcagctcaaaagaaagcagaag
     961 ---------+---------+---------+---------+---------+---------+ 1020
         cttctgggtttccgtcgttttgttctcagggagatttcgtcgagttttctttcgtcttc

a        E  D  P  K  A  A  K  T  R  V  P  L  K  Q  L  K  R  K  Q  K  -
b         K  T  Q  R  Q  Q  K  Q  E  S  L  *  S  S  S  K  E  S  R  S  -
c          R  P  K  G  S  K  N  K  S  P  S  K  A  A  Q  K  K  A  E  A -

         ccactggagaaaaacggccaagaggcagacctaggaaatggccacaacaagttgttcaga
    1021 ---------+---------+---------+---------+---------+---------+ 1080
         ggtgacctctttttgccggttctccgtctggatcctttaccggtgttgttcaacaagtct

a        P  L  E  K  N  G  Q  E  A  D  L  G  N  G  H  N  K  L  F  R  -
b         H  W  R  K  T  A  K  R  Q  T  *  E  M  A  T  T  S  C  S  E  -
c          T  G  E  K  R  P  R  G  R  P  R  K  W  P  Q  Q  V  V  Q  K -

         agaagcctgctcaggaggaaactgaagagacatcctcacaagagtctgccgaagaggact
    1081 ---------+---------+---------+---------+---------+---------+ 1140
         tcttcggacgagtcctcctttgacttctctgtaggagtgttctcagacggcttctcctga

a        R  S  L  L  R  R  K  L  K  R  H  P  H  K  S  L  P  K  R  T  -
b         E  A  C  S  G  G  N  *  R  D  I  L  T  R  V  C  R  R  G  L  -
c          K  P  A  Q  E  E  T  E  E  T  S  S  Q  E  S  A  E  E  D  * -

         agggggcgccaacgttcgatttctacctcagcagcagttggatcttttgaagggagaaga
    1141 ---------+---------+---------+---------+---------+---------+ 1200
         tcccccgcggttgcaagctaaagatggagtcgtcgtcaacctagaaaacttccctcttct

a        R  G  R  Q  R  S  I  S  T  S  A  A  V  G  S  F  E  G  R  R  -
b         G  G  A  N  V  R  F  L  P  Q  Q  Q  L  D  L  L  K  G  E     -
c          G  A  P  T  F  D  F  Y  L  S  S  S  W  I  F  *  R  E  K   -
```

Figure 19-1 NSP-A

(Linear) MAP of: hsnspa check: 4619 from: 1 to: 3202

RL;HSNSPA - Homo sapiens neuroendocrine-specific protein A (NSP) mRNA, complete
ID HSNSPA standard; RNA; HUM; 3202 BP.
AC L10333;
NI g307306
DT 16-JUN-1993 (Rel. 36, Created)
DT 18-JAN-1995 (Rel. 42, Last updated, Version 2) . . .

With 1 enzymes: NOTI

```
          ctgagacaccgcagcttccctgagcgccgagtccctccggggacagcagcagggagcgcc
        1 ---------+---------+---------+---------+---------+---------+ 60
          gactctgtggcgtcgaagggactcgcggctcagggaggcccctgtcgtcgtccctcgcgg

a         L  R  H  R  S  F  P  E  R  R  V  P  P  G  T  A  A  G  S  A  -
b          *  D  T  A  A  S  L  S  A  E  S  L  R  G  Q  Q  Q  G  A  P  -
c           E  T  P  Q  L  P  *  A  P  S  P  S  G  D  S  S  R  E  R  P -

          cgcgcagccaccgagcctctgcccagccaagccgccgtcgccgcgccgggggaccgccag
       61 ---------+---------+---------+---------+---------+---------+ 120
          gcgcgtcggtggctcggagacgggtcggttcggcggcagcggcgcggcccccctggcggtc

a         R  A  A  T  E  P  L  P  S  Q  A  A  V  A  A  P  G  D  R  Q  -
b          A  Q  P  P  S  L  C  P  A  K  P  P  S  P  R  R  G  T  A  S  -
c           R  S  H  R  A  S  A  Q  P  S  R  R  R  R  A  G  G  P  P  A -

          ccatggccgcgccgggggatccgcaggacgagctgctgccgctggccggccccgggtccc
      121 ---------+---------+---------+---------+---------+---------+ 180
          ggtaccggcgcggcccccctaggcgtcctgctcgacgacggcgaccggccggggcccaggg

a         P  W  P  R  R  G  I  R  R  T  S  C  C  R  W  P  A  P  G  P  -
b          H  G  R  A  G  G  S  A  G  R  A  A  A  A  G  R  P  R  V  P  -
c           M  A  A  P  G  D  P  Q  D  E  L  L  P  L  A  G  P  G  S  Q -

          agtggctcaggcaccgaggggaggggagaacgaagcggtgacgccgaaaggggccacgc
      181 ---------+---------+---------+---------+---------+---------+ 240
          tcaccgagtccgtggctcccctcccccctcttgcttcgccactgcggctttccccggtgcg

a         S  G  S  G  T  E  G  R  G  R  T  K  R  *  R  R  K  G  P  R  -
b          V  A  Q  A  P  R  G  G  G  E  R  S  G  D  A  E  R  G  H  A  -
c           W  L  R  H  R  G  E  G  E  N  E  A  V  T  P  K  G  A  T  P -
```

Figure 19-2

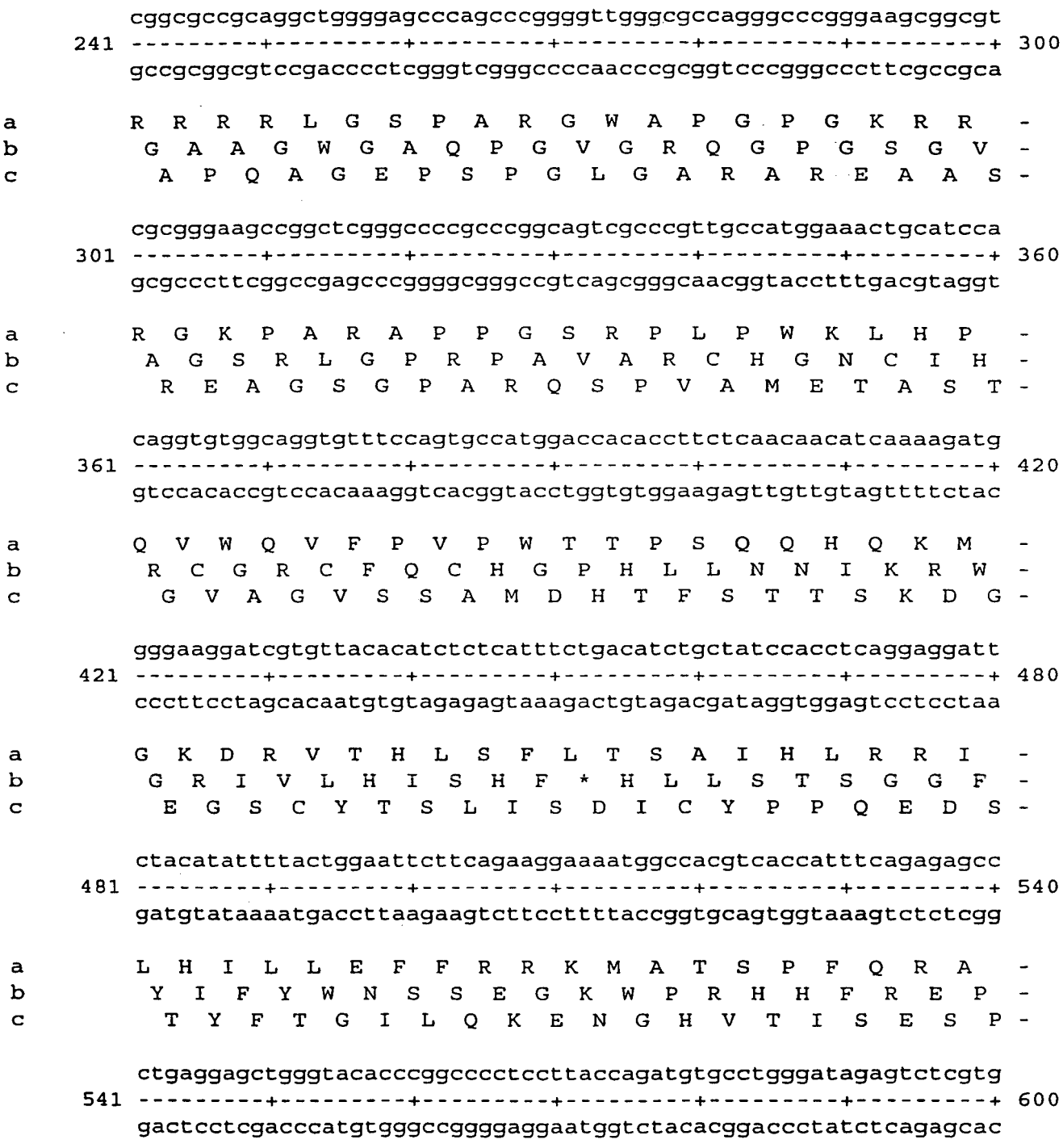
cggcgccgcaggctggggagcccagcccggggttgggcgccagggcccgggaagcggcgt
241 ---------+---------+---------+---------+---------+---------+---------+ 300
gccgcggcgtccgacccctcgggtcgggccccaacccgcggtcccgggcccttcgccgca
a R R R R L G S P A R G W A P G P G K R R -
b G A A G W G A Q P G V G R Q G P G S G V -
c A P Q A G E P S P G L G A R A R E A A S -
cgcgggaagccggctcgggccccgcccggcagtcgcccgttgccatggaaactgcatcca
301 ---------+---------+---------+---------+---------+---------+---------+ 360
gcgcccttcggccgagcccggggcgggccgtcagcgggcaacggtacctttgacgtaggt
a R G K P A R A P P G S R P L P W K L H P -
b A G S R L G P R P A V A R C H G N C I H -
c R E A G S G P A R Q S P V A M E T A S T -
caggtgtggcaggtgtttccagtgccatggaccacaccttctcaacaacatcaaaagatg
361 ---------+---------+---------+---------+---------+---------+---------+ 420
gtccacaccgtccacaaaggtcacggtacctggtgtggaagagttgttgtagttttctac
a Q V W Q V F P V P W T T P S Q Q H Q K M -
b R C G R C F Q C H G P H L L N N I K R W -
c G V A G V S S A M D H T F S T T S K D G -
gggaaggatcgtgttacacatctctcatttctgacatctgctatccacctcaggaggatt
421 ---------+---------+---------+---------+---------+---------+---------+ 480
cccttcctagcacaatgtgtagagagtaaagactgtagacgataggtggagtcctcctaa
a G K D R V T H L S F L T S A I H L R R I -
b G R I V L H I S H F * H L L S T S G G F -
c E G S C Y T S L I S D I C Y P P Q E D S -
ctacatattttactggaattcttcagaaggaaaatggccacgtcaccatttcagagagcc
481 ---------+---------+---------+---------+---------+---------+---------+ 540
gatgtataaaatgaccttaagaagtcttcctttaccggtgcagtggtaaagtctctcgg
a L H I L L E F F R R K M A T S P F Q R A -
b Y I F Y W N S S E G K W P R H H F R E P -
c T Y F T G I L Q K E N G H V T I S E S P -
ctgaggagctgggtacacccggcccctccttaccagatgtgcctgggatagagtctcgtg
541 ---------+---------+---------+---------+---------+---------+---------+ 600
gactcctcgacccatgtgggccggggaggaatggtctacacggaccctatctcagagcac

Figure 19-3

```
a       L  R  S  W  V  H  P  A  P  P  Y  Q  M  C  L  G  *  S  L  V  -
b        *  G  A  G  Y  T  R  P  L  L  T  R  C  A  W  D  R  V  S  W  -
c         E  E  L  G  T  P  G  P  S  L  P  D  V  P  G  I  E  S  R  G -

         gcttatttagttctgattctggaatagagatgactcctgcagagtccacggaagtgaaca
    601  ---------+---------+---------+---------+---------+---------+ 660
         cgaataaatcaagactaagaccttatctctactgaggacgtctcaggtgccttcacttgt

a       A  Y  L  V  L  I  L  E  *  R  *  L  L  Q  S  P  R  K  *  T  -
b        L  I  *  F  *  F  W  N  R  D  D  S  C  R  V  H  G  S  E  Q  -
c         L  F  S  S  D  S  G  I  E  M  T  P  A  E  S  T  E  V  N  K -

         agatcttagcagaccctctggaccagatgaaagcagaggcctataaatacattgacataa
    661  ---------+---------+---------+---------+---------+---------+ 720
         tctagaatcgtctgggagacctggtctactttcgtctccggatatttatgtaactgtatt

a       R  S  *  Q  T  L  W  T  R  *  K  Q  R  P  I  N  T  L  T  *  -
b        D  L  S  R  P  S  G  P  D  E  S  R  G  L  *  I  H  *  H  N  -
c         I  L  A  D  P  L  D  Q  M  K  A  E  A  Y  K  Y  I  D  I  T -

         ccagacccgaggaggtgaagcaccaagaacaacatcaccccgagctggaagataaagact
    721  ---------+---------+---------+---------+---------+---------+ 780
         ggtctgggctcctccacttcgtggttcttgttgtagtggggctcgaccttctatttctga

a       P  D  P  R  R  *  S  T  K  N  N  I  T  P  S  W  K  I  K  T  -
b        Q  T  R  G  G  E  A  P  R  T  T  S  P  R  A  G  R  *  R  L  -
c         R  P  E  E  V  K  H  Q  E  Q  H  H  P  E  L  E  D  K  D  L -

         tggactttaagaataaagacactgacatctcaattaaacctgaaggagtccgtgaacctg
    781  ---------+---------+---------+---------+---------+---------+ 840
         acctgaaattcttatttctgtgactgtagagttaatttggacttcctcaggcacttggac

a       W  T  L  R  I  K  T  L  T  S  Q  L  N  L  K  E  S  V  N  L  -
b        G  L  *  E  *  R  H  *  H  L  N  *  T  *  R  S  P  *  T  *  -
c         D  F  K  N  K  D  T  D  I  S  I  K  P  E  G  V  R  E  P  D -

         acaaaccagctcctgtggagggaaaaatcatcaaggaccatttattggaagaatccacat
    841  ---------+---------+---------+---------+---------+---------+ 900
         tgtttggtcgaggacacctccctttttagtagttcctggtaaataaccttcttaggtgta

a       T  N  Q  L  L  W  R  E  K  S  S  R  T  I  Y  W  K  N  P  H  -
b        Q  T  S  S  C  G  G  K  N  H  Q  G  P  F  I  G  R  I  H  I  -
c         K  P  A  P  V  E  G  K  I  I  K  D  H  L  L  E  E  S  T  F -
```

Figure 19-4

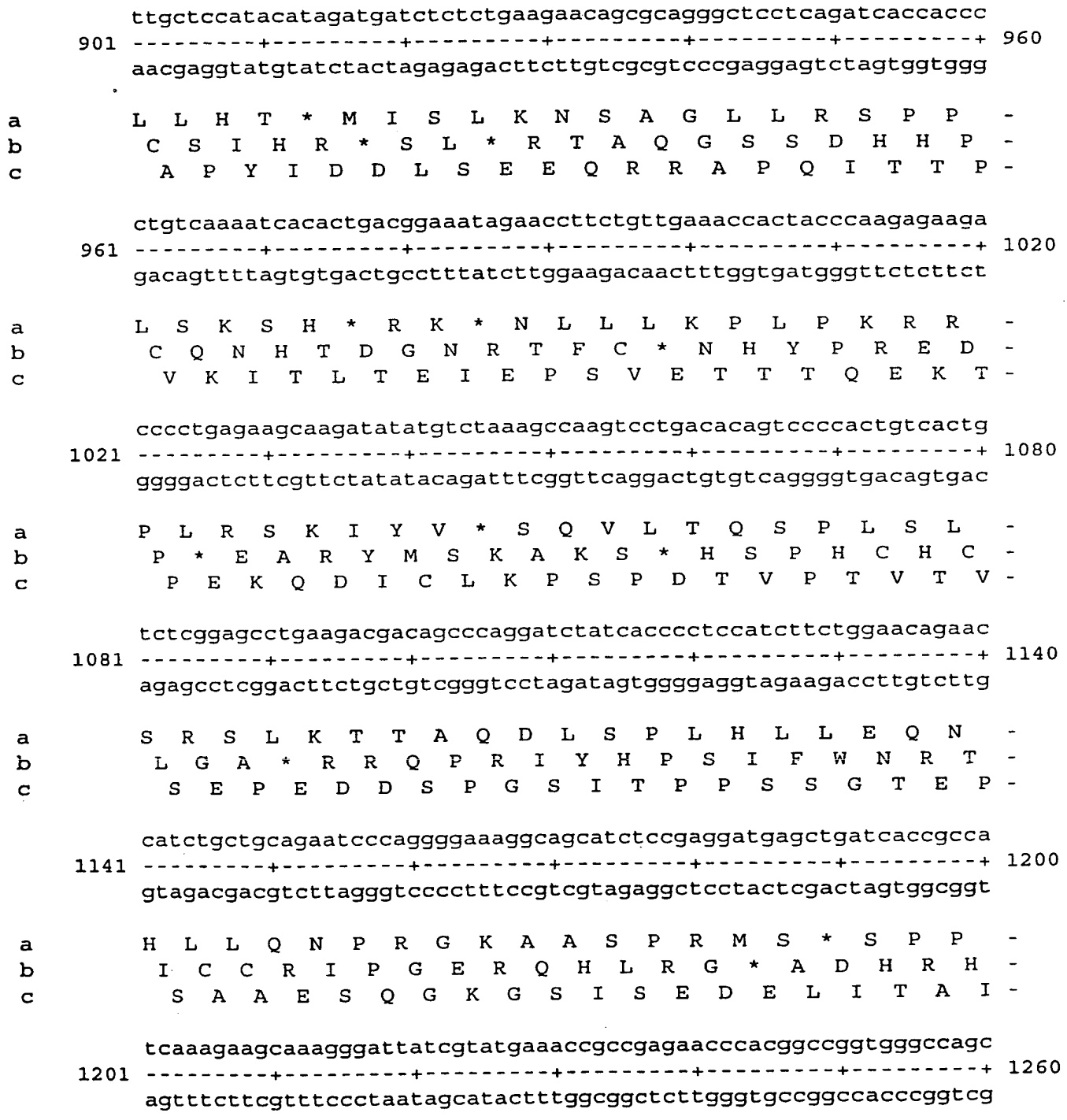
ttgctccatacatagatgatctctctgaagaacagcgcagggctcctcagatcaccaccc
901 ---------+---------+---------+---------+---------+---------+---------+ 960
aacgaggtatgtatctactagagagacttcttgtcgcgtcccgaggagtctagtggtggg
a L L H T * M I S L K N S A G L L R S P P -
b C S I H R * S L * R T A Q G S S D H H P -
c A P Y I D D L S E E Q R R A P Q I T T P -
ctgtcaaaatcacactgacggaaatagaaccttctgttgaaaccactacccaagagaaga
961 ---------+---------+---------+---------+---------+---------+---------+ 1020
gacagttttagtgtgactgcctttatcttggaagacaactttggtgatgggttctcttct
a L S K S H * R K * N L L L K P L P K R R -
b C Q N H T D G N R T F C * N H Y P R E D -
c V K I T L T E I E P S V E T T T Q E K T -
cccctgagaagcaagatatatgtctaaagccaagtcctgacacagtccccactgtcactg
1021 ---------+---------+---------+---------+---------+---------+---------+ 1080
ggggactcttcgttctatatacagatttcggttcaggactgtgtcaggggtgacagtgac
a P L R S K I Y V * S Q V L T Q S P L S L -
b P * E A R Y M S K A K S * H S P H C H C -
c P E K Q D I C L K P S P D T V P T V T V -
tctcggagcctgaagacgacagcccaggatctatcacccctccatcttctggaacagaac
1081 ---------+---------+---------+---------+---------+---------+---------+ 1140
agagcctcggacttctgctgtcgggtcctagatagtggggaggtagaagaccttgtcttg
a S R S L K T T A Q D L S P L H L L E Q N -
b L G A * R R Q P R I Y H P S I F W N R T -
c S E P E D D S P G S I T P P S S G T E P -
catctgctgcagaatcccaggggaaaggcagcatctccgaggatgagctgatcaccgcca
1141 ---------+---------+---------+---------+---------+---------+---------+ 1200
gtagacgacgtcttagggtccccttccgtcgtagaggctcctactcgactagtggcggt
a H L L Q N P R G K A A S P R M S * S P P -
b I C C R I P G E R Q H L R G * A D H R H -
c S A A E S Q G K G S I S E D E L I T A I -
tcaaagaagcaaagggattatcgtatgaaaccgccgagaacccacggccggtgggccagc
1201 ---------+---------+---------+---------+---------+---------+---------+ 1260
agtttcttcgtttccctaatagcatactttggcggctcttgggtgccggccacccggtcg

Figure 19-5

```
a        S  K  K  Q  R  D  Y  R  M  K  P  P  R  T  H  G  R  W  A  S   -
b         Q  R  S  K  G  I  I  V  *  N  R  R  E  P  T  A  G  G  P  A  -
c          K  E  A  K  G  L  S  Y  E  T  A  E  N  P  R  P  V  G  Q  L -

         tggccgacaggcccgaggtcaaggccaggtccggaccgccaaccatccccagccccctgg
    1261 ---------+---------+---------+---------+---------+---------+ 1320
         accggctgtccgggctccagttccggtccaggcctggcggttggtaggggtcgggggacc

a        W  P  T  G  P  R  S  R  P  G  P  D  R  Q  P  S  P  A  P  W   -
b         G  R  Q  A  R  G  Q  G  Q  V  R  T  A  N  H  P  Q  P  P  G  -
c          A  D  R  P  E  V  K  A  R  S  G  P  P  T  I  P  S  P  L  D -

         accacgaggccagcagcgcggagtcgggggactcagagatcgagctggtgtccgaggacc
    1321 ---------+---------+---------+---------+---------+---------+ 1380
         tggtgctccggtcgtcgcgcctcagcccccctgagtctctagctcgaccacaggctcctgg

a        T  T  R  P  A  A  R  S  R  G  T  Q  R  S  S  W  C  P  R  T   -
b         P  R  G  Q  Q  R  G  V  G  G  L  R  D  R  A  G  V  R  G  P  -
c          H  E  A  S  S  A  E  S  G  D  S  E  I  E  L  V  S  E  D  P -

         ccatggccgcggaggacgcgctgccctcaggctatgtgagctttggccacgtgggcggcc
    1381 ---------+---------+---------+---------+---------+---------+ 1440
         ggtaccggcgcctcctgcgcgacgggagtccgatacactcgaaaccggtgcacccgccgg

a        P  W  P  R  R  T  R  C  P  Q  A  M  *  A  L  A  T  W  A  A   -
b         H  G  R  G  G  R  A  A  L  R  L  C  E  L  W  P  R  G  R  P  -
c          M  A  A  E  D  A  L  P  S  G  Y  V  S  F  G  H  V  G  G  P -

         cgccgccctcgcccgcctcgccatccatccagtacagcatcctgagggaggagcgcgagg
    1441 ---------+---------+---------+---------+---------+---------+ 1500
         gcggcgggagcgggcggagcggtaggtaggtcatgtcgtaggactccctcctcgcgctcc

a        R  R  P  R  P  P  R  H  P  S  S  T  A  S  *  G  R  S  A  R   -
b         A  A  L  A  R  L  A  I  H  P  V  Q  H  P  E  G  G  A  R  G  -
c          P  P  S  P  A  S  P  S  I  Q  Y  S  I  L  R  E  E  R  E  A -

         ccgagctggacagcgagctcatcatcgagtcgtgcgacgcctcctcggcctcggaggaga
    1501 ---------+---------+---------+---------+---------+---------+ 1560
         ggctcgacctgtcgctcgagtagtagctcagcacgctgcggaggagccggagcctcctct

a        P  S  W  T  A  S  S  S  S  S  R  A  T  P  P  R  P  R  R  R   -
b         R  A  G  Q  R  A  H  H  R  V  V  R  R  L  L  G  L  G  G  E  -
c          E  L  D  S  E  L  I  I  E  S  C  D  A  S  S  A  S  E  E  S -
```

Figure 19-6

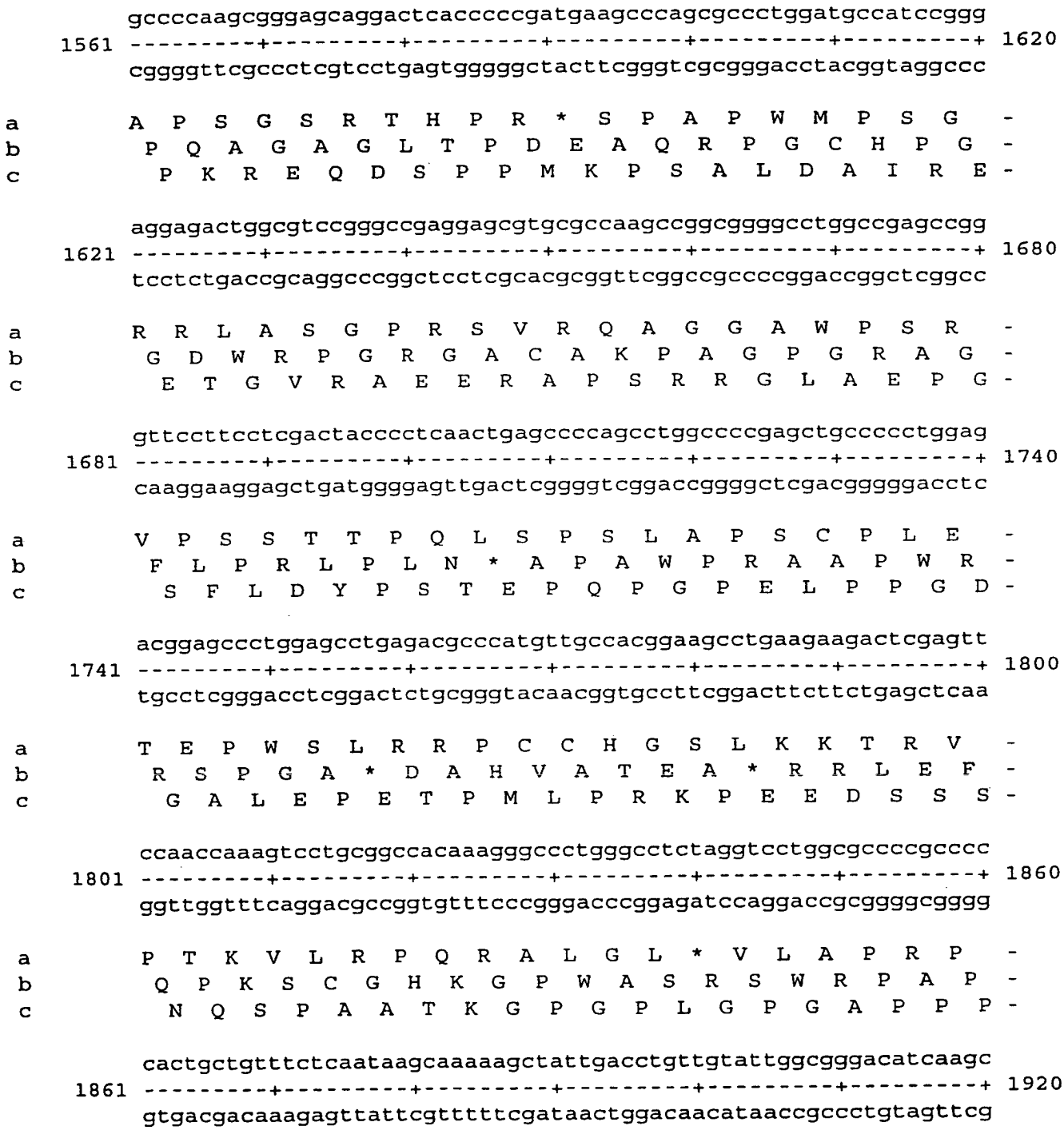
gccccaagcgggagcaggactcacccccgatgaagcccagcgccctggatgccatccggg
1561 ---------+---------+---------+---------+---------+---------+ 1620
cggggttcgccctcgtcctgagtgggggctacttcgggtcgcgggacctacggtaggccc
a A P S G S R T H P R * S P A P W M P S G -
b P Q A G A G L T P D E A Q R P G C H P G -
c P K R E Q D S P P M K P S A L D A I R E -
aggagactggcgtccgggccgaggagcgtgcgccaagccggcggggcctggccgagccgg
1621 ---------+---------+---------+---------+---------+---------+ 1680
tcctctgaccgcaggcccggctcctcgcacgcggttcggccgccccggaccggctcggcc
a R R L A S G P R S V R Q A G G A W P S R -
b G D W R P G R G A C A K P A G P G R A G -
c E T G V R A E E R A P S R R G L A E P G -
gttccttcctcgactacccctcaactgagccccagcctggccccgagctgccccctggag
1681 ---------+---------+---------+---------+---------+---------+ 1740
caaggaaggagctgatggggagttgactcggggtcggaccggggctcgacgggggacctc
a V P S S T T P Q L S P S L A P S C P L E -
b F L P R L P L N * A P A W P R A A P W R -
c S F L D Y P S T E P Q P G P E L P P G D -
acggagccctggagcctgagacgcccatgttgccacggaagcctgaagaagactcgagtt
1741 ---------+---------+---------+---------+---------+---------+ 1800
tgcctcgggacctcggactctgcgggtacaacggtgccttcggacttcttctgagctcaa
a T E P W S L R R P C C H G S L K K T R V -
b R S P G A * D A H V A T E A * R R L E F -
c G A L E P E T P M L P R K P E E D S S S -
ccaaccaaagtcctgcggccacaaagggccctgggcctctaggtcctggcgccccgcccc
1801 ---------+---------+---------+---------+---------+---------+ 1860
ggttggtttcaggacgccggtgtttcccgggacccggagatccaggaccgcggggcgggg
a P T K V L R P Q R A L G L * V L A P R P -
b Q P K S C G H K G P W A S R S W R P A P -
c N Q S P A A T K G P G P L G P G A P P P -
cactgctgtttctcaataagcaaaaagctattgacctgttgtattggcgggacatcaagc
1861 ---------+---------+---------+---------+---------+---------+ 1920
gtgacgacaaagagttattcgtttttcgataactggacaacataaccgccctgtagttcg

Figure 19-7

```
a        H  C  C  F  S  I  S  K  K  L  L  T  C  C  I  G  G  T  S  S  -
b         T  A  V  S  Q  *  A  K  S  Y  *  P  V  V  L  A  G  H  Q  A  -
c          L  L  F  L  N  K  Q  K  A  I  D  L  L  Y  W  R  D  I  K  Q -

         agacgggcatcgtgtttgggagtttcctgctgctgctcttctccctgacccagttcagcg
    1921 ---------+---------+---------+---------+---------+---------+ 1980
         tctgcccgtagcacaaaccctcaaaggacgacgacgagaagagggactgggtcaagtcgc

a        R  R  A  S  C  L  G  V  S  C  C  C  S  S  P  *  P  S  S  A  -
b         D  G  H  R  V  W  E  F  P  A  A  A  L  L  P  D  P  V  Q  R  -
c          T  G  I  V  F  G  S  F  L  L  L  L  F  S  L  T  Q  F  S  V -

         tggtgagcgtcgtggcctacctggccctggccgcactctcagccaccatcagtttccgca
    1981 ---------+---------+---------+---------+---------+---------+ 2040
         accactcgcagcaccggatggaccgggaccggcgtgagagtcggtggtagtcaaaggcgt

a        W  *  A  S  W  P  T  W  P  W  P  H  S  Q  P  P  S  V  S  A  -
b         G  E  R  R  G  L  P  G  P  G  R  T  L  S  H  H  Q  F  P  H  -
c          V  S  V  V  A  Y  L  A  L  A  A  L  S  A  T  I  S  F  R  I -

         tctacaagtctgttttacaagcagtgcagaaaaccgacgaaggccacccttttcaaggcct
    2041 ---------+---------+---------+---------+---------+---------+ 2100
         agatgttcagacaaaatgttcgtcacgtcttttggctgcttccggtgggaaagttccgga

a        S  T  S  L  F  Y  K  Q  C  R  K  P  T  K  A  T  L  S  R  P  -
b         L  Q  V  C  F  T  S  S  A  E  N  R  R  R  P  P  F  Q  G  L  -
c          Y  K  S  V  L  Q  A  V  Q  K  T  D  E  G  H  P  F  K  A  Y -

         acttggagcttgagatcacccttttctcaggagcagattcagaagtacacggactgcctgc
    2101 ---------+---------+---------+---------+---------+---------+ 2160
         tgaacctcgaactctagtgggaaagagtcctcgtctaagtcttcatgtgcctgacggacg

a        T  W  S  L  R  S  P  F  L  R  S  R  F  R  S  T  R  T  A  C  -
b         L  G  A  *  D  H  P  F  S  G  A  D  S  E  V  H  G  L  P  A  -
c          L  E  L  E  I  T  L  S  Q  E  Q  I  Q  K  Y  T  D  C  L  Q -

         agttctacgtgaacagcacacttaaggaactgaggaggctcttccttgtccaggacctgg
    2161 ---------+---------+---------+---------+---------+---------+ 2220
         tcaagatgcacttgtcgtgtgaattccttgactcctccgagaaggaacaggtcctggacc

a        S  S  T  *  T  A  H  L  R  N  *  G  G  S  S  L  S  R  T  W  -
b         V  L  R  E  Q  H  T  *  G  T  E  E  A  L  P  C  P  G  P  G  -
c          F  Y  V  N  S  T  L  K  E  L  R  R  L  F  L  V  Q  D  L  V -
```

Figure 19-8

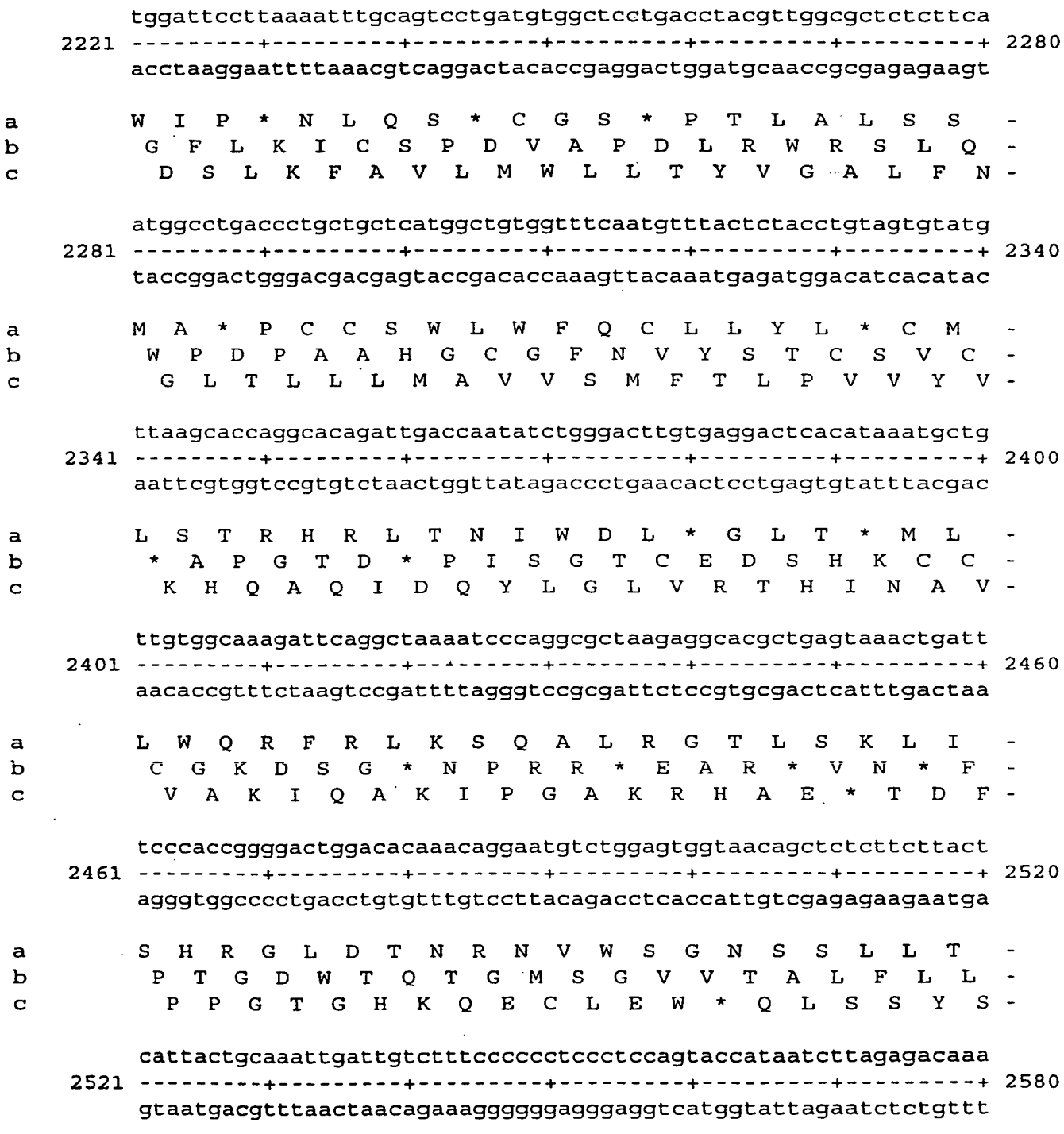
tggattccttaaaatttgcagtcctgatgtggctcctgacctacgttggcgctctcttca
2221 2280
acctaaggaattttaaacgtcaggactacaccgaggactggatgcaaccgcgagagaagt
a W I P * N L Q S * C G S * P T L A L S S -
b G F L K I C S P D V A P D L R W R S L Q -
c D S L K F A V L M W L L T Y V G A L F N -
atggcctgaccctgctgctcatggctgtggtttcaatgtttactctacctgtagtgtatg
2281 2340
taccggactgggacgacgagtaccgacaccaaagttacaaatgagatggacatcacatac
a M A * P C C S W L W F Q C L L Y L * C M -
b W P D P A A H G C G F N V Y S T C S V C -
c G L T L L L M A V V S M F T L P V V Y V -
ttaagcaccaggcacagattgaccaatatctgggacttgtgaggactcacataaatgctg
2341 2400
aattcgtggtccgtgtctaactggttatagaccctgaacactcctgagtgtatttacgac
a L S T R H R L T N I W D L * G L T * M L -
b * A P G T D * P I S G T C E D S H K C C -
c K H Q A Q I D Q Y L G L V R T H I N A V -
ttgtggcaaagattcaggctaaaatcccaggcgctaagaggcacgctgagtaaactgatt
2401 2460
aacaccgtttctaagtccgattttagggtccgcgattctccgtgcgactcatttgactaa
a L W Q R F R L K S Q A L R G T L S K L I -
b C G K D S G * N P R R * E A R * V N * F -
c V A K I Q A K I P G A K R H A E * T D F -
tcccaccggggactggacacaaacaggaatgtctggagtggtaacagctctcttcttact
2461 2520
agggtggcccctgacctgtgtttgtccttacagacctcaccattgtcgagagaagaatga
a S H R G L D T N R N V W S G N S S L L T -
b P T G D W T Q T G M S G V V T A L F L L -
c P P G T G H K Q E C L E W * Q L S S Y S -
cattactgcaaattgattgtctttccccctccctccagtaccataatcttagagacaaa
2521 2580
gtaatgacgtttaactaacagaaaggggggagggaggtcatggtattagaatctctgttt

Figure 19-9

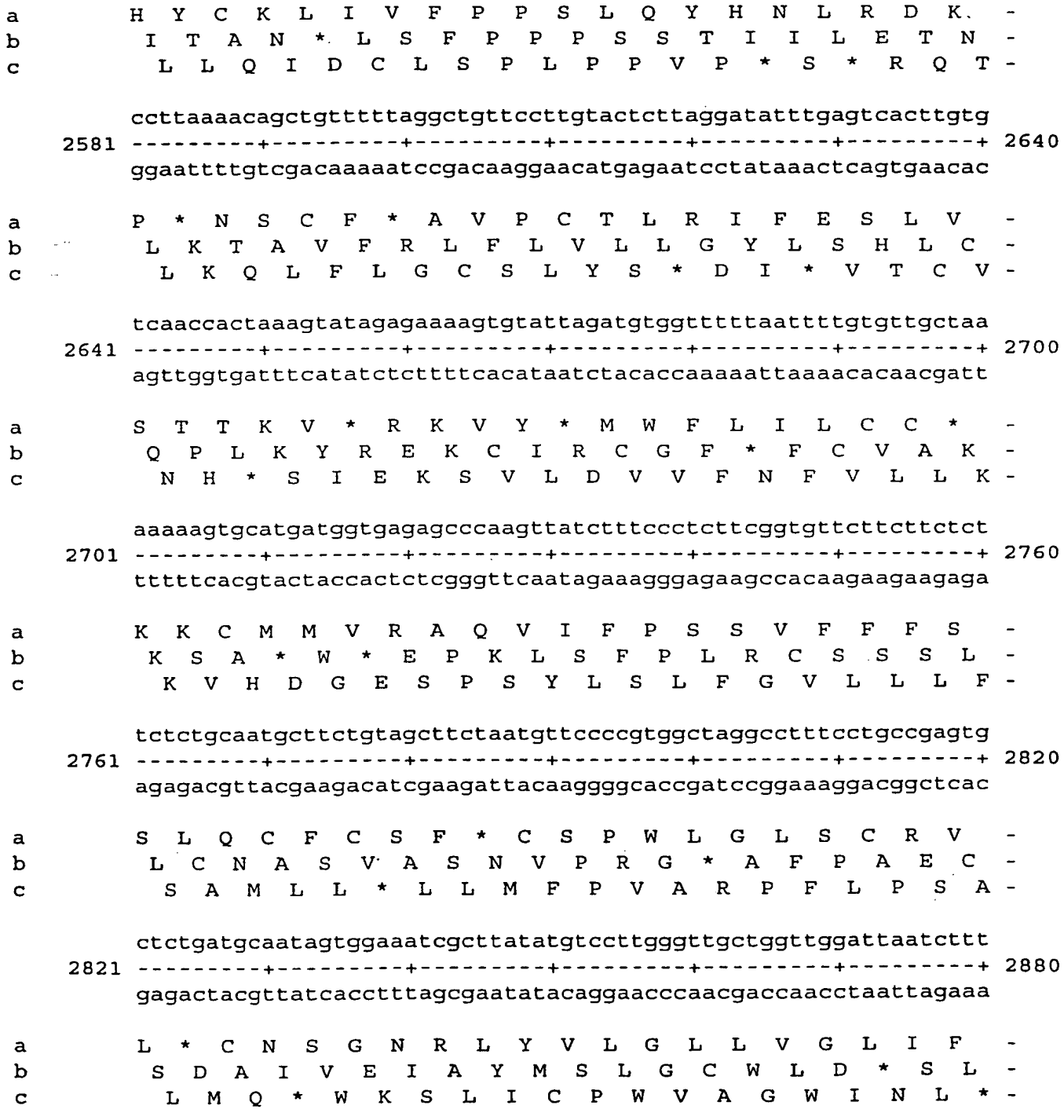
a H Y C K L I V F P P S L Q Y H N L R D K -
b I T A N * L S F P P P S S T I I L E T N -
c L L Q I D C L S P L P P V P * S * R Q T -
ccttaaaacagctgtttttaggctgttccttgtactcttaggatatttgagtcacttgtg
2581 ---------+---------+---------+---------+---------+---------+ 2640
ggaattttgtcgacaaaaatccgacaaggaacatgagaatcctataaactcagtgaacac
a P * N S C F * A V P C T L R I F E S L V -
b L K T A V F R L F L V L L G Y L S H L C -
c L K Q L F L G C S L Y S * D I * V T C V -
tcaaccactaaagtatagagaaaagtgtattagatgtggtttttaattttgtgttgctaa
2641 ---------+---------+---------+---------+---------+---------+ 2700
agttggtgatttcatatctcttttcacataatctacaccaaaaattaaaacacaacgatt
a S T T K V * R K V Y * M W F L I L C C * -
b Q P L K Y R E K C I R C G F * F C V A K -
c N H * S I E K S V L D V V F N F V L L K -
aaaaagtgcatgatggtgagagcccaagttatctttccctcttcggtgttcttcttctct
2701 ---------+---------+---------+---------+---------+---------+ 2760
tttttcacgtactaccactctcgggttcaatagaaagggagaagccacaagaagaagaga
a K K C M M V R A Q V I F P S S V F F F S -
b K S A * W * E P K L S F P L R C S S S L -
c K V H D G E S P S Y L S L F G V L L L F -
tctctgcaatgcttctgtagcttctaatgttccccgtggctaggcctttcctgccgagtg
2761 ---------+---------+---------+---------+---------+---------+ 2820
agagacgttacgaagacatcgaagattacaaggggcaccgatccggaaaggacggctcac
a S L Q C F C S F * C S P W L G L S C R V -
b L C N A S V A S N V P R G * A F P A E C -
c S A M L L * L L M F P V A R P F L P S A -
ctctgatgcaatagtggaaatcgcttatatgtccttgggttgctggttggattaatcttt
2821 ---------+---------+---------+---------+---------+---------+ 2880
gagactacgttatcacctttagcgaatatacaggaacccaacgaccaacctaattagaaa
a L * C N S G N R L Y V L G L L V G L I F -
b S D A I V E I A Y M S L G C W L D * S L -
c L M Q * W K S L I C P W V A G W I N L * -

Figure 19-10

```
      aataacaatatatagaattgtagactgatgttttagcatttttccaacacacacaacgta
 2881 ---------+---------+---------+---------+---------+---------+ 2940
      ttattgttatatatcttaacatctgactacaaaatcgtaaaaaggttgtgtgtgttgcat

a         N  N  N  I  *  N  C  R  L  M  F  *  H  F  S  N  T  H  N  V  -
b          I  T  I  Y  R  I  V  D  *  C  F  S  I  F  P  T  H  T  T  *  -
c           *  Q  Y  I  E  L  *  T  D  V  L  A  F  F  Q  H  T  Q  R  K -

      aaaataaaagcagtcgaccgcacttatggtaatcagttttgtataacttaaaataattaa
 2941 ---------+---------+---------+---------+---------+---------+ 3000
      ttttattttcgtcagctggcgtgaataccattagtcaaaacatattgaattttattaatt

a         K  I  K  A  V  D  R  T  Y  G  N  Q  F  C  I  T  *  N  N  *  -
b          K  *  K  Q  S  T  A  L  M  V  I  S  F  V  *  L  K  I  I  K  -
c           N  K  S  S  R  P  H  L  W  *  S  V  L  Y  N  L  K  *  L  N -

      ataaatgaataaatccaaaacaaacatgcagtacttttgttgtatgggattggtgggctg
 3001 ---------+---------+---------+---------+---------+---------+ 3060
      tatttacttatttaggttttgtttgtacgtcatgaaaacaacataccctaaccacccgac

a         I  N  E  *  I  Q  N  K  H  A  V  L  L  L  Y  G  I  G  G  L  -
b          *  M  N  K  S  K  T  N  M  Q  Y  F  C  C  M  G  L  V  G  *  -
c           K  *  I  N  P  K  Q  T  C  S  T  F  V  V  W  D  W  W  A  D -

      atttacatgtatggttactaaaaagtaccagcatgttaactttattacaatttgtattac
 3061 ---------+---------+---------+---------+---------+---------+ 3120
      taaatgtacataccaatgatttttcatggtcgtacaattgaaataatgttaaacataatg

a         I  Y  M  Y  G  Y  *  K  V  P  A  C  *  L  Y  Y  N  L  Y  Y  -
b          F  T  C  M  V  T  K  K  Y  Q  H  V  N  F  I  T  I  C  I  T  -
c           L  H  V  W  L  L  K  S  T  S  M  L  T  L  L  Q  F  V  L  L -

      tttctctgtagttcctaatggattcaattacggactctggatatttgcacttatgtactt
 3121 ---------+---------+---------+---------+---------+---------+ 3180
      aaagagacatcaaggattacctaagttaatgcctgagacctataaacgtgaatacatgaa

a         F  L  C  S  S  *  W  I  Q  L  R  T  L  D  I  C  T  Y  V  L  -
b          F  S  V  V  P  N  G  F  N  Y  G  L  W  I  F  A  L  M  Y  L  -
c           S  L  *  F  L  M  D  S  I  T  D  S  G  Y  L  H  L  C  T  * -

      gatactgaatgcataaataaat
 3181 ---------+----------+-- 3202
      ctatgacttacgtatttattta
```

Figure 19-11

```
a        D  T  E  C  I  N  K    -
b         I  L  N  A  *  I  N   -
c          Y  *  M  H  K  *     -
```

Enzymes that do cut:

NONE

Enzymes that do not cut:

NotI

Figure 20

3 amyloid precursor protein[1] (exons 9 and 10):

Sequence	Type	SEQ ID
GAGAGGCTTGAGGCCAAGCACCGAGAGAGAATGTCCCAGGTCATGAGAGAATGGGAAGAGGCAGAACGTCAAGCAAAGAACTTGCCTAAA	wt nucleotides	[SEQ ID NO: 25]
E R L E A K H R E R M S Q V M R E W E E A E R Q A K N L P K	wt protein	[SEQ ID NO: 83]
E R L E A K H R E N V P G H E R M G R G R T S S K E L A *	+1 protein	[SEQ ID NO: 1388]
GAGAGGCTTGAGGCCAAGCACCGAGAGAATGTCCCAGGTCATGAGAGAATGGGAAGAGGCAGAACGTCAAGCAAAGAACTTGCCTAAA (ΔGA)	+1 nucleotides	[SEQ ID NO: 1389]
E R L E A K H R E R M S Q V M R M G R G R T S S K E L A *	+1 protein	[SEQ ID NO: 1390]
GAGAGGCTTGAGGCCAAGCACCGAGAGAGAATGTCCCAGGTCATGAGAATGGGAAGAGGCAGAACGTCAAGCAAAGAACTTGCCTAAA (ΔGA)	+1 nucleotides	[SEQ ID NO: 1391]

Ubiquitin B[2] (exon 2, repeat 1/2):
2nd repeat

Sequence	Type	SEQ ID
CACCTGGTCCTGCGTCTGAGAGGTGGTATGCAGATCTTCGTGAAGACCCTGACCGGCAAGACCATCACCCTGGAAGTGGAGCCCAGTGAC	wt nucleotides	[SEQ ID NO: 113]
H L V L R L R G G M Q I F V K T L T G K T I T L E V E P S D	wt protein	[SEQ ID NO: 125]
H L V L R L R G Y A D L R E D P D R Q D H H P G S G A Q *	+1 protein	[SEQ ID NO: 1392]
CACCTGGTCCTGCGTCTGAGAGGGTATGCAGATCTTCGTGAAGACCCTGACCGGCAAGACCATCACCCTGGAAGTGGAGCCCAGTGAC (ΔGT)	+1 nucleotides	[SEQ ID NO: 1393]

Ubiquitin B (exon 2, repeat 2/3):
3rd repeat

Sequence	Type	SEQ ID
CACCTGGTCCTGCGTCTGAGAGGTGGTATGCAGATCTTCGTGAAGACCCTGACCGGCAAGACCATCACTCTGGAGGTGGAGCCCAGTGAC	wt nucleotides	[SEQ ID NO: 113]
H L V L R L R G G M Q I F V K T L T G K T I T L E V E P S D	wt protein	[SEQ ID NO: 125]
H L V L R L R G G M Q I F V K T L T G K T I T G G G A Q *	+1 protein	[SEQ ID NO: 1394]
CACCTGGTCCTGCGTCTGAGAGGTGGTATGCAGATCTTCGTGAAGACCCTGACCGGCAAGACCATCACTGGAGGTGGAGCCCAGTGAC (ΔCT)	+1 nucleotide	[SEQ ID NO: 1395]

[1]Number of GAGAG motifs: 7. Predicted molecular weight of truncated protein 38 kDa [30].
[2]Number of GAGAG motifs: 2. Predicted molecular weight of truncated protein 11 KDa (monomer) and expressed in brain [31,32].
▾ exon 9/10 junction

Figure 21-1 Restriction sites generated by dinucleotide deletions in transcripts of ß amyloid precursor protein (βAPP) and Ubiquitin B

In general, a mutation in the nucleotide sequence can result in changes in the restriction enzyme recognition sites of the sequence. The GA deletion in exon 9 of ß-APP and the GT deletion in the first repeat of of Ubi-B do not alter the restriction map of the sequence down- and upstream the deletion. However, due to the GA deletion in exon 10 of ß-APP an Msl-I site is created at the site of the deletion (Fig. 21-2,-3,-4).
In the Ubi-B sequence the CT deletion in the second repeat leads to the loss of a Hin4-I and BstX-I site and the creation of a Cje-I site upstream and the creation of a BsR-I and a TspR-I site downstream the deletion site (Fig. 21-5-8).

Figure 21-2

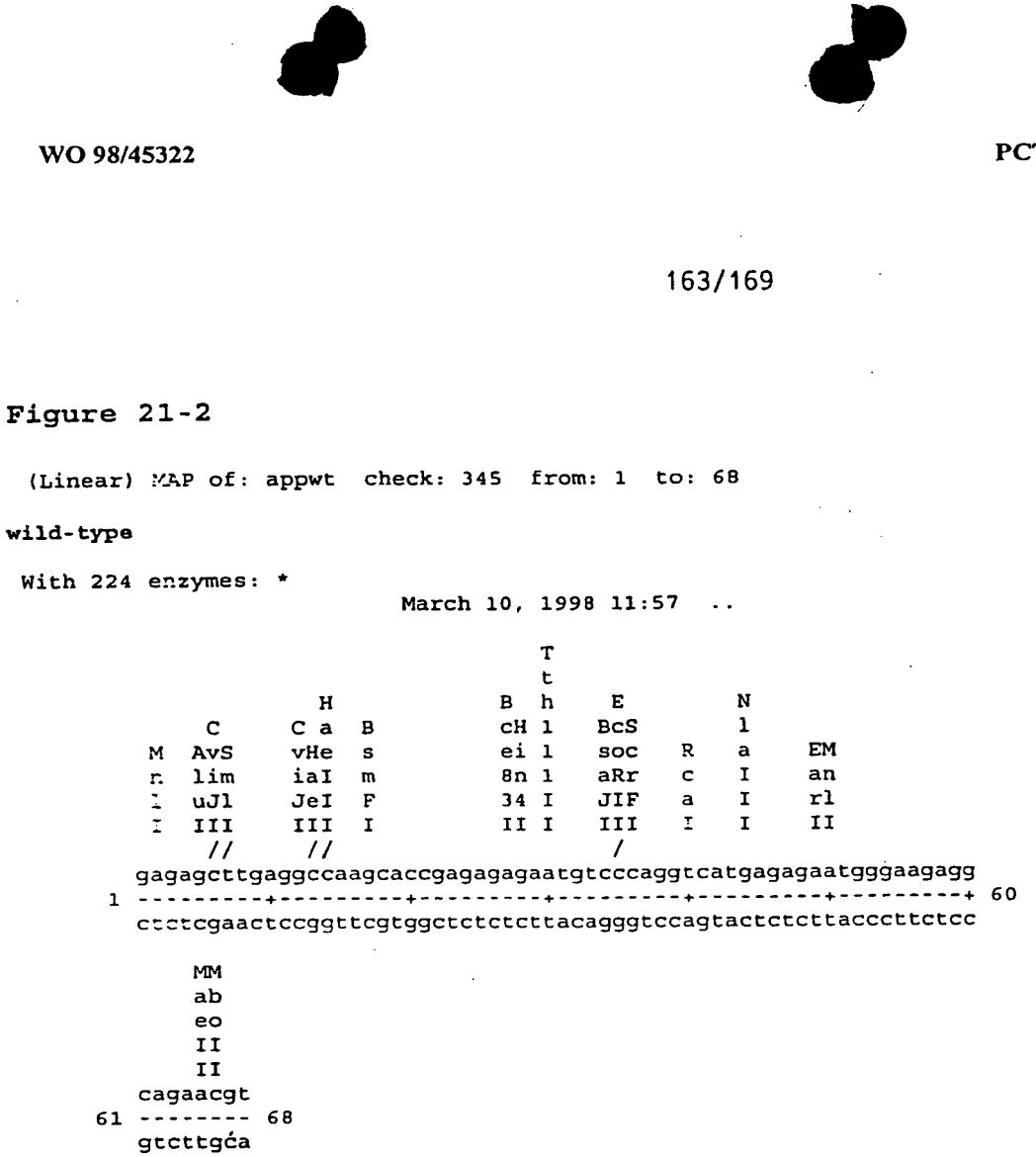
(Linear) MAP of: appwt check: 345 from: 1 to: 68
wild-type
With 224 enzymes: *
March 10, 1998 11:57 ..
gagagcttgaggccaagcaccgagagagaatgtcccaggtcatgagagaatgggaagagg
ctctcgaactccggttcgtggctctctcttacagggtccagtactctcttacccttctcc
cagaacgt
gtcttgca
Enzymes that do cut:
AluI Bce83I BsaJI BsmFI CviJI EarI EcoRII HaeI
HaeIII Hin4I MaeII MboII MnlI NlaIII RcaI ScrFI
SmlI Tth111II
Enzymes that do not cut:
AarI AatII AccI AceIII AciI AclI AflII AflIII
AhdI AloI AlwI AlwNI ApaI ApaLI ApoI AscI
AvaI AvaII AvrII BaeI BaeI BamHI BanI BanII
BbsI BbvI BbvCI BccI BcefI BcgI BcgI BciVI
BclI BfaI BglI BglII BmgI BmrI BplI BplI
BpmI Bpu10I Bpu1102I BsaI BsaAI BsaBI BsaHI BsaWI
BsaXI BsbI BscGI BseMII BseRI BseSI BsgI BsiEI
BsiHKAI BslI BsmI BsmAI BsmBI Bsp24I Bsp24I Bsp1286I
BspEI BspGI BspLU11I BspMI BsrI BsrBI BsrDI BsrFI
BsrGI BssHII BssSI Bst4CI BstAPI BstDSI BstEII BstXI
BstYI BstZ17I Bsu36I BtsI Cac8I CjeI CjeI CjePI
CjePI ClaI CviRI DdeI DpnI DraI DraIII DrdI
DrdII EaeI EagI EciI Eco47III Eco57I EcoNI Eco0109I
EcoRI EcoRV FauI Fnu4HI FokI FseI FspI GdiII
HaeII HaeIV HgaI HgiEII HhaI HincII HindIII HinfI
HpaI HphI KpnI MaeIII MluI MmeI MscI MseI
MslI MspI MspA1I MunI MwoI NarI NciI NcoI
NdeI NgoAIV NheI NlaIV NotI NruI NsiI NspI
NspV PacI Pfl1108I PflMI PinAI PleI PmeI PmlI
PshAI Psp5II PstI PvuI PvuII RleAI RsaI RsrII
SacI SacII SalI SanDI SapI Sau3AI Sau96I SbfI
ScaI SexAI SfaNI SfcI SfiI SgfI SgrAI SimI
SmaI SnaBI SpeI SphI SrfI Sse8647I SspI Sth132I
StuI StyI SunI SwaI TaqI TaqII TaqII TatI
TauI TfiI ThaI TseI Tsp45I Tsp509I TspRI Tth111I
UbaLI VspI XbaI XcmI XhoI XmnI

Figure 21-3

(Linear) MAP of: **appex9 ΔGA** check: 8734 from: 1 to: 66

GA-deletie exon 9: no difference in restriction map with wild type (Fig. 21-2)

With 224 enzymes: *

March 10, 1998 11:57 ..

```
                                    T
                                    t
                  H                 B  h  E          N
        C         C aB              Hc  1 BcS        1
    M  AvS        vHes              ie  1 soc    R   a     EM
    n  lim        iaIm              n8  1 aRr    c   I     an
    1  uJi        JeIF              43  I JIF    a   I     rl
    I  III        IIII              II  I III    I   I     II
        //        //                      /
  gagagccttgaggccaagcaccgagagaatgtcccaggtcatgagagaatgggaagaggca
1 ---------+---------+---------+---------+---------+---------+ 60
  ctctcgaactccggttcgtggctctcttacagggtccagtactctcttacccttctccgt

    MM
    ab
    eo
    II
    II
  gaacgt
61 ------ 66
  cttgca
```

Enzymes that do cut:

AluI	Bce83I	BsaJI	BsmFI	CviJI	EarI	EcoRII	HaeI
HaeIII	Hin4I	MaeII	MboII	MnlI	NlaIII	RcaI	ScrFI
SmlI	Tth111II						

Enzymes that do not cut:

AarI	AatII	AccI	AceIII	AciI	AclI	AflII	AflIII
AhdI	AloI	AlwI	AlwNI	ApaI	ApaLI	ApoI	AscI
AvaI	AvaII	AvrII	BaeI	BaeI	BamHI	BanI	BanII
BbsI	BbvI	BbvCI	BccI	BcefI	BcgI	BcgI	BciVI
BclI	BfaI	BglI	BglII	BmgI	BmrI	BplI	BplI
BpmI	Bpu10I	Bpu1102I	BsaI	BsaAI	BsaBI	BsaHI	BsaWI
BsaXI	BsbI	BscGI	BseMII	BseRI	BseSI	BsgI	BsiEI
BsiHKAI	BslI	BsmI	BsmAI	BsmBI	Bsp24I	Bsp24I	Bsp1286I
BspEI	BspGI	BspLU11I	BspMI	BsrI	BsrBI	BsrDI	BsrFI
BsrGI	BssHII	BssSI	Bst4CI	BstAPI	BstDSI	BstEII	BstXI
BstYI	BstZ17I	Bsu36I	BtsI	Cac8I	CjeI	CjeI	CjePI
CjePI	ClaI	CviRI	DdeI	DpnI	DraI	DraIII	DrdI
DrdII	EaeI	EagI	EciI	Eco47III	Eco57I	EcoNI	Eco0109I
EcoRI	EcoRV	FauI	Fnu4HI	FokI	FseI	FspI	GdiII
HaeII	HaeIV	HgaI	HgiEII	HhaI	HincII	HindIII	HinfI
HpaI	HphI	KpnI	MaeIII	MluI	MmeI	MscI	MseI
MslI	MspI	MspA1I	MunI	MwoI	NarI	NciI	NcoI
NdeI	NgoAIV	NheI	NlaIV	NotI	NruI	NsiI	NspI
NspV	PacI	Pfl1108I	PflMI	PinAI	PleI	PmeI	PmlI
PshAI	Psp5II	PstI	PvuI	PvuII	RleAI	RsaI	RsrII
SacI	SacII	SalI	SanDI	SapI	Sau3AI	Sau96I	SbfI
ScaI	SexAI	SfaNI	SfcI	SfiI	SgfI	SgrAI	SimI
SmaI	SnaBI	SpeI	SphI	SrfI	Sse8647I	SspI	Sth132I
StuI	StyI	SunI	SwaI	TaqI	TaqII	TaqII	TatI
TauI	TfiI	ThaI	TseI	Tsp45I	Tsp509I	TspRI	Tth111I
UbaLI	VspI	XbaI	XcmI	XhoI	XmnI		

Figure 21-4

(Linear) MAP of: appex10 ΔGA check: 8854 from: 1 to: 66

GA-deletie exon 10: no difference in restriction map with wild type (Fig. 21-2)

With 224 enzymes: *

March 10, 1998 11:57 ..

```
                                    T
                                    t
                  H                 B  h    E          N
          C       C a  B            cH 1    BcS        l
      M   AvS     vHe  s            ei 1    soc    R   aM EM
      n   lim     iaI  m            8n 1    aRr    c   Is an
      l   uJl     JeI  F            34 I    JIF    a   Il rl
      I   III     III  I            II I    III    I   II II
          //      //                      /
    gagagcttgaggccaagcaccgagagagaatgtcccaggtcatgagaatgggaagaggca
  1 ---------+---------+---------+---------+---------+---------+ 60
    ctctcgaactccggttcgtggctctctcttacagggtccagtactcttacccttctccgt

      MM
      ab
      eo
      II
      II
    gaacgt
 61 ------ 66
    cttgca
```

Enzymes that do cut:

AluI	Bce83I	BsaJI	BsmFI	CviJI	EarI	EcoRII	HaeI
HaeIII	Hin4I	MaeII	MboII	MnlI	MslI	NlaIII	RcaI
ScrFI	SmlI	Tth111II					

Enzymes that do not cut:

AarI	AatII	AccI	AceIII	AciI	AclI	AflII	AflIII
AhdI	AloI	AlwI	AlwNI	ApaI	ApaLI	ApoI	AscI
AvaI	AvaII	AvrII	BaeI	BaeI	BamHI	BanI	BanII
BbsI	BbvI	BbvCI	BccI	BcefI	BcgI	BcgI	BciVI
BclI	BfaI	BglI	BglII	BmgI	BmrI	BplI	BplI
BpmI	Bpu10I	Bpu1102I	BsaI	BsaAI	BsaBI	BsaHI	BsaWI
BsaXI	BsbI	BscGI	BseMII	BseRI	BseSI	BsgI	BsiEI
BsiHKAI	BslI	BsmI	BsmAI	BsmBI	Bsp24I	Bsp24I	Bsp1286I
BspEI	BspGI	BspLU11I	BspMI	BsrI	BsrBI	BsrDI	BsrFI
BsrGI	BssHII	BssSI	Bst4CI	BstAPI	BstDSI	BstEII	BstXI
BstYI	BstZ17I	Bsu36I	BtsI	Cac8I	CjeI	CjeI	CjePI
CjePI	ClaI	CviRI	DdeI	DpnI	DraI	DraIII	DrdI
DrdII	EaeI	EagI	EciI	Eco47III	Eco57I	EcoNI	Eco0109I
EcoRI	EcoRV	FauI	Fnu4HI	FokI	FseI	FspI	GdiII
HaeII	HaeIV	HgaI	HgiEII	HhaI	HincII	HindIII	HinfI
HpaI	HphI	KpnI	MaeIII	MluI	MmeI	MscI	MseI
MspI	MspA1I	MunI	MwoI	NarI	NciI	NcoI	NdeI
NgoAIV	NheI	NlaIV	NotI	NruI	NsiI	NspI	NspV
PacI	Pfl1108I	PflMI	PinAI	PleI	PmeI	PmlI	PshAI
Psp5II	PstI	PvuI	PvuII	RleAI	RsaI	RsrII	SacI
SacII	SalI	SanDI	SapI	Sau3AI	Sau96I	SbfI	ScaI
SexAI	SfaNI	SfcI	SfiI	SgfI	SgrAI	SimI	SmaI
SnaBI	SpeI	SphI	SrfI	Sse8647I	SspI	Sth132I	StuI
StyI	SunI	SwaI	TaqI	TaqII	TaqII	TatI	TauI
TfiI	ThaI	TseI	Tsp45I	Tsp509I	TspRI	Tth111I	UbaLI
VspI	XbaI	XcmI	XhoI	XmnI			

Figure 21-5

(Linear) MAP of: **ubiwt1** check: 3896 from: 1 to: 34

With 224 enzymes: *

March 10, 1998 11:07 ..

```
              B                  S
              s          M  C BB a
          M  De          b  v gsDu
          n  dM          o  i ltp3
          l  eI          I  R IYnA
          I  II          I  I IIII
                            / /
     tcctgcgtctgagaggtggtatgcagatcttcgt
   1 ---------+---------+---------+---- 34
     aggacgcagactctccaccatacgtctagaagca
```

Enzymes that do cut:

BglII	BseMII	BstYI	CviRI	DdeI	DpnI	MboII	MnlI
Sau3AI							

Enzymes that do not cut:

AarI	AatII	AccI	AceIII	AciI	AclI	AflII	AflIII
AhdI	AloI	AluI	AlwI	AlwNI	ApaI	ApaLI	ApoI
AscI	AvaI	AvaII	AvrII	BaeI	BaeI	BamHI	BanI
BanII	BbsI	BbvI	BbvCI	BccI	Bce83I	BcefI	BcgI
BcgI	BciVI	BclI	BfaI	BglI	BmgI	BmrI	BplI
BplI	BpmI	Bpu10I	Bpu1102I	BsaI	BsaAI	BsaBI	BsaHI
BsaJI	BsaWI	BsaXI	BsbI	BscGI	BseRI	BseSI	BsgI
BsiEI	BsiHKAI	BslI	BsmI	BsmAI	BsmBI	BsmFI	Bsp24I
Bsp24I	Bsp1286I	BspEI	BspGI	BspLU11I	BspMI	BsrI	BsrBI
BsrDI	BsrFI	BsrGI	BssHII	BssSI	Bst4CI	BstAPI	BstDSI
BstEII	BstXI	BstZ17I	Bsu36I	BtsI	Cac8I	CjeI	CjeI
CjePI	CjePI	ClaI	CviJI	DraI	DraIII	DrdI	DrdII
EaeI	EagI	EarI	EciI	Eco47III	Eco57I	EcoNI	Eco0109I
EcoRI	EcoRII	EcoRV	FauI	Fnu4HI	FokI	FseI	FspI
GdiII	HaeI	HaeII	HaeIII	HaeIV	HgaI	HgiEII	HhaI
Hin4I	HincII	HindIII	HinfI	HpaI	HphI	KpnI	MaeII
MaeIII	MluI	MmeI	MscI	MseI	MslI	MspI	MspA1I
MunI	MwoI	NarI	NciI	NcoI	NdeI	NgoAIV	NheI
NlaIII	NlaIV	NotI	NruI	NsiI	NspI	NspV	PacI
Pfl1108I	PflMI	PinAI	PleI	PmeI	PmlI	PshAI	Psp5II
PstI	PvuI	PvuII	RcaI	RleAI	RsaI	RsrII	SacI
SacII	SalI	SanDI	SapI	Sau96I	SbfI	ScaI	ScrFI
SexAI	SfaNI	SfcI	SfiI	SgfI	SgrAI	SimI	SmaI
SmlI	SnaBI	SpeI	SphI	SrfI	Sse8647I	SspI	Sth132I
StuI	StyI	SunI	SwaI	TaqI	TaqII	TaqII	TatI
TauI	TfiI	ThaI	TseI	Tsp45I	Tsp509I	TspRI	Tth111I
Tth111II	UbaLI	VspI	XbaI	XcmI	XhoI	XmnI	

Figure 21-6

(Linear) MAP of: ubi Δ gt check: 8822 from: 1 to: 32

With 224 enzymes: *

March 10, 1998 11:08 ..

```
              B                 S
              s         M   C BB a
          M   De        b   v gsDu
          n   dM        o   i ltp3
          l   eI        I   R IYnA
          I   II        I   I IIII
                              / /
     tcctgcgtctgagagggtatgcagatcttcgt
   1 ---------+---------+---------+-- 32
     aggacgcagactctcccatacgtctagaagca
```

Enzymes that do cut:

BglII	BseMII	BstYI	CviRI	DdeI	DpnI	MboII	MnlI
Sau3AI							

Enzymes that do not cut:

AarI	AatII	AccI	AceIII	AciI	AclI	AflII	AflIII
AhdI	AloI	AluI	AlwI	AlwNI	ApaI	ApaLI	ApoI
AscI	AvaI	AvaII	AvrII	BaeI	BaeI	BamHI	BanI
BanII	BbsI	BbvI	BbvCI	BccI	Bce83I	BcefI	BcgI
BcgI	BciVI	BclI	BfaI	BglI	BmgI	BmrI	BplI
BplI	BpmI	Bpu10I	Bpu1102I	BsaI	BsaAI	BsaBI	BsaHI
BsaJI	BsaWI	BsaXI	BsbI	BscGI	BseRI	BseSI	BsgI
BsiEI	BsiHKAI	BslI	BsmI	BsmAI	BsmBI	BsmFI	Bsp24I
Bsp24I	Bsp1286I	BspEI	BspGI	BspLU11I	BspMI	BsrI	BsrBI
BsrDI	BsrFI	BsrGI	BssHII	BssSI	Bst4CI	BstAPI	BstDSI
BstEII	BstXI	BstZ17I	Bsu36I	BtsI	Cac8I	CjeI	CjeI
CjePI	CjePI	ClaI	CviJI	DraI	DraIII	DrdI	DrdII
EaeI	EagI	EarI	EciI	Eco47III	Eco57I	EcoNI	Eco0109I
EcoRI	EcoRII	EcoRV	FauI	Fnu4HI	FokI	FseI	FspI
GdiII	HaeI	HaeII	HaeIII	HaeIV	HgaI	HgiEII	HhaI
Hin4I	HincII	HindIII	HinfI	HpaI	HphI	KpnI	MaeII
MaeIII	MluI	MmeI	MscI	MseI	MslI	MspI	MspA1I
MunI	MwoI	NarI	NciI	NcoI	NdeI	NgoAIV	NheI
NlaIII	NlaIV	NotI	NruI	NsiI	NspI	NspV	PacI
Pfl1108I	PflMI	PinAI	PleI	PmeI	PmlI	PshAI	Psp5II
PstI	PvuI	PvuII	RcaI	RleAI	RsaI	RsrII	SacI
SacII	SalI	SanDI	SapI	Sau96I	SbfI	ScaI	ScrFI
SexAI	SfaNI	SfcI	SfiI	SgfI	SgrAI	SimI	SmaI
SmlI	SnaBI	SpeI	SphI	SrfI	Sse8647I	SspI	Sth132I
StuI	StyI	SunI	SwaI	TaqI	TaqII	TaqII	TatI
TauI	TfiI	ThaI	TseI	Tsp45I	Tsp509I	TspRI	Tth111I
Tth111II	UbaLI	VspI	XbaI	XcmI	XhoI	XmnI	

Figure 21-7

```
(Linear) MAP of: ubiwt2  check: 8814  from: 1  to: 28

With 224 enzymes: *

                          March 10, 1998 11:08  ..

                 H    B
                Bi M  s
                cn n  t
                c4 l  X
                II I  I
      ggcaagaccatcactctggaggtggagc
    1 ---------+---------+-------- 28
      ccgttctggtagtgagacctccacctcg
```

Enzymes that do cut:

BccI	BstXI	Hin4I	MnlI

Enzymes that do not cut:

AarI	AatII	AccI	AceIII	AciI	AclI	AflII	AflIII
AhdI	AloI	AluI	AlwI	AlwNI	ApaI	ApaLI	ApoI
AscI	AvaI	AvaII	AvrII	BaeI	BaeI	BamHI	BanI
BanII	BbsI	BbvI	BbvCI	Bce83I	BcefI	BcgI	BcgI
BciVI	BclI	BfaI	BglI	BglII	BmgI	BmrI	BplI
BplI	BpmI	Bpu10I	Bpu1102I	BsaI	BsaAI	BsaBI	BsaHI
BsaJI	BsaWI	BsaXI	BsbI	BscGI	BseMII	BseRI	BseSI
BsgI	BsiEI	BsiHKAI	BslI	BsmI	BsmAI	BsmBI	BsmFI
Bsp24I	Bsp24I	Bsp1286I	BspEI	BspGI	BspLU11I	BspMI	BsrI
BsrBI	BsrDI	BsrFI	BsrGI	BssHII	BssSI	Bst4CI	BstAPI
BstDSI	BstEII	BstYI	BstZ17I	Bsu36I	BtsI	Cac8I	CjeI
CjeI	CjePI	CjePI	ClaI	CviJI	CviRI	DdeI	DpnI
DraI	DraIII	DrdI	DrdII	EaeI	EagI	EarI	EciI
Eco47III	Eco57I	EcoNI	Eco0109I	EcoRI	EcoRII	EcoRV	FauI
Fnu4HI	FokI	FseI	FspI	GdiII	HaeI	HaeII	HaeIII
HaeIV	HgaI	HgiEII	HhaI	HincII	HindIII	HinfI	HpaI
HphI	KpnI	MaeII	MaeIII	MboII	MluI	MmeI	MscI
MseI	MslI	MspI	MspAlI	MunI	MwoI	NarI	NciI
NcoI	NdeI	NgoAIV	NheI	NlaIII	NlaIV	NotI	NruI
NsiI	NspI	NspV	PacI	Pfl1108I	PflMI	PinAI	PleI
PmeI	PmlI	PshAI	Psp5II	PstI	PvuI	PvuII	RcaI
RleAI	RsaI	RsrII	SacI	SacII	SalI	SanDI	SapI
Sau3AI	Sau96I	SbfI	ScaI	ScrFI	SexAI	SfaNI	SfcI
SfiI	SgfI	SgrAI	SimI	SmaI	SmlI	SnaBI	SpeI
SphI	SrfI	Sse8647I	SspI	Sth132I	StuI	StyI	SunI
SwaI	TaqI	TaqII	TaqII	TatI	TauI	TfiI	ThaI
TseI	Tsp45I	Tsp509I	TspRI	Tth111I	Tth111II	UbaLI	VspI
XbaI	XcmI	XhoI	XmnI				

Figure 21-8

```
(Linear) MAP of: ubi Δ ct  check: 4758  from: 1  to: 26

With 224 enzymes: *

                          March 10, 1998 11:08  ..

                                 T
          C      BM            Bs
          j      cn            sp
          e      cl            rR
          I      II            II
                                /
        ggcaagaccatcactggaggtggagc
      1 ---------+---------+------ 26
        ccgttctggtagtgacctccacctcg

Enzymes that do cut:

    BccI      BsrI      CjeI      MnlI     TspRI

Enzymes that do not cut:

    AarI     AatII      AccI    AceIII      AciI      AclI     AflII    AflIII
    AhdI      AloI      AluI      AlwI     AlwNI      ApaI     ApaLI      ApoI
    AscI      AvaI     AvaII     AvrII      BaeI      BaeI     BamHI      BanI
   BanII      BbsI      BbvI     BbvCI    Bce83I     BcefI      BcgI      BcgI
   BciVI      BclI      BfaI      BglI     BglII      BmgI      BmrI      BplI
    BplI      BpmI    Bpu10I   Bpu1102I      BsaI     BsaAI     BsaBI     BsaHI
   BsaJI     BsaWI     BsaXI      BsbI     BscGI    BseMII     BseRI     BseSI
    BsgI      BsiEI   BsiHKAI      BslI      BsmI     BsmAI     BsmBI     BsmFI
  Bsp24I    Bsp24I   Bsp1286I     BspEI     BspGI   BspLU11I     BspMI     BsrBI
   BsrDI     BsrFI     BsrGI     BssHII     BssSI     Bst4CI     BstAPI    BstDSI
  BstEII     BstXI     BstYI    BstZ17I    Bsu36I      BtsI      Cac8I      CjePI
   CjePI      ClaI     CviJI      CviRI      DdeI      DpnI      DraI      DraIII
    DrdI     DrdII      EaeI      EagI      EarI      EciI    Eco47III    Eco57I
   EcoNI   Eco0109I     EcoRI     EcoRII     EcoRV      FauI     Fnu4HI      FokI
    FseI      FspI      GdiII      HaeI     HaeII    HaeIII     HaeIV      HgaI
   HgiEII     HhaI      Hin4I     HincII   HindIII     HinfI      HpaI      HphI
    KpnI     MaeII     MaeIII     MboII      MluI      MmeI      MscI      MseI
    MslI      MspI     MspA1I      MunI      MwoI      NarI      NciI      NcoI
    NdeI    NgoAIV      NheI     NlaIII     NlaIV      NotI      NruI      NsiI
    NspI      NspV      PacI   Pfl1108I     PflMI     PinAI      PleI      PmeI
    PmlI     PshAI    Psp5II      PstI      PvuI      PvuII      RcaI     RleAI
    RsaI     RsrII      SacI      SacII      SalI      SanDI      SapI    Sau3AI
  Sau96I      SbfI      ScaI      ScrFI     SexAI      SfaNI      SfcI      SfiI
    SgfI     SgrAI      SimI      SmaI      SmlI      SnaBI      SpeI      SphI
    SrfI   Sse8647I      SspI    Sth132I      StuI      StyI      SunI      SwaI
    TaqI     TaqII     TaqII      TatI      TauI      TfiI      ThaI      TseI
  Tsp45I   Tsp509I   Tth111I   Tth111II     UbaLI     VspI      XbaI      XcmI
    XhoI      XmnI
```